

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2005, 16:58:20 ; Search time 5102 Seconds
(without alignments)
12641.773 Million cell updates/sec

Title: US-09-750-986D-29

Perfect score: 1770

Sequence: 1 atcgctctctcgagcgct.....gtgacgagcgacactgtga 1770

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST.*
1: gb_ests1.*
2: gb_ests2.*
3: gb_ests3.*
4: gb_ests4.*
5: gb_ests5.*
6: gb_ests6.*
7: gb_ests7.*
8: gb_ests8.*
9: gb_ests9.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	159.4	9.0	708	8	AZ934119 BJ Ba000
C 2	139.8	7.9	671	8	AZ933866 BJ Ba000
C 3	109.6	6.2	544	8	AZ933488 BJ Ba000
C 4	99.2	5.6	714	8	AZ933744 BJ Ba000
C 5	47.4	2.7	1101	9	CNS0175Y
C 6	43.4	2.5	613	6	CD872800
C 7	42.8	2.4	660	7	CK350537
C 8	42.6	2.4	585	5	BX321259
C 9	42.6	2.4	795	5	BX882027
C 10	42.6	2.4	815	5	BX911066
C 11	41.4	2.3	839	4	CNS004NB
C 12	41.2	2.3	240	4	BM442982
C 13	41.2	2.3	363	4	BM443022
C 14	40.8	2.3	856	4	CR676208
C 15	40.8	2.3	873	3	CR662673
C 16	40.8	2.3	875	3	CR673735
C 17	40.8	2.3	883	3	CR641761
C 18	40.8	2.3	890	3	CR654791
C 19	40.8	2.3	893	3	CR674838
C 20	40.8	2.3	902	3	CR638062
C 21	40.8	2.3	908	3	CR690234
C 22	40.8	2.3	1200	3	CR660267
C 23	40.8	2.3	1209	3	CR662146
C 24	40.8	2.3	1214	3	CR639857

25	40.6	2.3	1176	8	B2551949
C 26	40.2	2.3	863	5	B0221883
C 27	40.2	2.3	992	6	CD557447
28	39.8	2.2	646	5	BX296590
29	39.4	2.2	453	7	CR614906
30	39.4	2.2	735	5	BU311454
31	39.4	2.2	776	7	CO422920
C 32	39.2	2.2	456	3	CR686551
C 33	39.2	2.2	467	9	CU197417
C 34	39.2	2.2	508	1	AU187981
C 35	39.2	2.2	577	3	CR709583
C 36	39.2	2.2	623	3	CR642382
C 37	39.2	2.2	628	6	CB924736
C 38	39.2	2.2	839	3	CR678116
C 39	39.2	2.2	861	3	CR659440
C 40	39.2	2.2	873	3	CR674191
C 41	39.2	2.2	873	3	CR688437
C 42	39.2	2.2	874	3	CR677284
C 43	39.2	2.2	876	3	CR658925
C 44	39.2	2.2	876	3	CR676220
C 45	39.2	2.2	877	3	CR703088

ALIGNMENTS

RESULT 1
AZ934119/c 708 bp DNA linear GSS 24-APR-2001
LOCUS BJ_Ba0001H18f B. japonicum BAC library Bradyrhizobium japonicum
DEFINITION genomic, genomic survey sequence.
ACCESSION AZ934119
VERSION AZ934119.1 GI:13776179
KEYWORDS GSS.
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
REFERENCE Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A., Goicoechea, J.L., Stacey, G., Sadowsky, M.J. and Wing, R.A.
A marker-dense, sequence-ready map of the Bradyrhizobium japonicum genome
Genome Res. 11 (8), 1434-1440 (2001)
JOURNAL MEDLINE 21376150
PUBMED 11483585
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Class: BAC ends
High quality sequence stop: 661.
FEATURES
source
location/Qualifiers
1..708
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/note="Vector: pindig0536; Site_1: HindIII"
ORIGIN
Query Match 9.0%; Score 159.4; DB 8; Length 708;
Best local similarity 53.8%; Pred. No. 3.9e-33;
Matches 328; Conservative 0; Mismatches 282; Indels 0; Gaps 0;
QY 650 CTTTCCCGTTTGTGTAAGACCCGCGTCTGTGTAAGTGTGCGGTGTAAGACCA 709
DB 614 CTTTCTTACAGGCGTGAACGATCTGTGTAATCTTGACTGCGTCCGTGACCA 555

[illegible]

Source	1. .671	/organism="Bradyrhizobium japonicum"
		/mol_type="genomic DNA"
		/strain="USD110"
		/db_xref="taxon:375"
		/lab_host="E. coli"
		/clone_id="B. japonicum BAC library"
		/note="Vector: pindig0536; Site_1: HindIII"

Query Match	7.9%	Score 139.8	DB 8	Length 671
Best Local Similarity	52.7%	Pred. No. 1.2e-27		
Matches 323	Conservative 0	Mismatches 289	Indels 1	Gaps 1

LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM
A2935488	544 bp DNA linear	GSS	24-APR-2001		
A2935488	BT_Bac0003J24r B. japonicum BAC library				
A2935488	genomic, genomic survey sequence.				
A2935488	GI:13778288				
GSS.					
Bradyrhizobium japonicum					
Bradyrhizobium japonicum					
Bacteria: Proteobacteria; Alphaproteobacteria; Rhizobiales;					
Bradyrhizobiaceae; Bradyrhizobium.					
1 (bases 1 to 544)					
Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,					

TITLE
A marker-dense, sequence-ready map of the *Bradyrhizobium japonicum*
Genome Res. 11 (8), 1434-1440 (2001)
JOURNAL
MEDLINE
PUBMED
11483585
COMMENT
Contact: Ming RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Class: BAC ends
High quality sequence stop: 499.
Location/Qualifiers

FEATURES
source
1. .544
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"
/strain="USD110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/note="Vector: pindig0536; Site_1: HindIII"

ORIGIN

Query Match 6.2%; Score 109.6; DB 8; Length 544;
Best Local Similarity 52.4%; Pred. No. 3.3e-19;
Matches 241; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 1273 GATCTGCGCATCTCTGAGAAAGGTGATGTTGAGACGATGCTGAAAGACTTCAAG 1332
DB 2 GATAGCGCGCGCGCGCGCGCGCGCTTTCTTGCACGCGCGGTGTCGCGAGACTTCAAG 61
QY 1333 CTGCTCTCAAGGGGATTTGTCAAGCGTTGGCCATTTGCGCAACGCGCGGCTTCTGAAAGC 1392
DB 62 CTCATTTCCGGACCTGCGGTCAAGCGTCAAGCGTCAAGCGTCAAGCGTCAAGCGTCAAG 121
QY 1393 GCGCTTTAGCTCTGACGATGAGTGTGCTGCTCTGATGATGATGATGATGATGATGATG 1452
DB 122 GCGCGCTCGCGAGGATATGTCGTGCGCGGTATGCGCGCGCGCGCGCGCGCGCTTCTG 181
QY 1453 GTGTTTCCGCGCTCTCTGCACTGCGCGCTTGTGCGCGCGCTGAGAAAGAGCGCTCGAC 1512
DB 182 GTGTTTCCGCGCATCTCTGCGGTGCGCGCGCTTGTGCGCGCGCTTGTGCGCGCGCTG 241
QY 1513 GCGAGGTGCTTGCAGATGAGCGCGGTGCGCGCGCTTGTGCTGCTGCTGCTGCTGCTGCTG 1572
DB 242 AATGATGTGCTGCGACATGACAGATCAAGGAGTGCATGCTGCTGCTGCTGCTGCTGCTG 301
QY 1573 AATCGAAGCAACTGCGCAATGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1632
DB 302 AACGAGAGGCGCGCAATTCCTCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTG 361
QY 1633 CCGTCAATGATTAAGGCGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1692
DB 362 CCAATCGGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTG 421
QY 1693 CAATGCGGTGCGCGAAGATTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1732
DB 422 ACGCGGCGCGCGAGGCGGTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 461

RESULT 4
A2933744/c 714 bp DNA linear GSS 24-Apr-2001
LOCUS
DEFINITION
Bj Ba0001A14r B. japonicum BAC library Bradyrhizobium japonicum
ACCESSION
A2933744
VERSION
A2933744.1 GI:13775804
KEYWORDS
GSS.
SOURCE
Bradyrhizobium japonicum
ORGANISM
Bradyrhizobium japonicum

REFERENCE
Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A.,
Goicoechea, J.L., Stacey, G., Sadowsky, M.J., and Ming, R.A.,
A marker-dense, sequence-ready map of the *Bradyrhizobium japonicum*
Genome Res. 11 (8), 1434-1440 (2001)
JOURNAL
MEDLINE
PUBMED
11483585
COMMENT
Contact: Ming RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Class: BAC ends
High quality sequence start: 23
High quality sequence stop: 664.
Location/Qualifiers

FEATURES

source
1. .714
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"
/strain="USD110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/note="Vector: pindig0536; Site_1: HindIII"

ORIGIN

Query Match 5.6%; Score 99.2; DB 8; Length 714;
Best Local Similarity 48.9%; Pred. No. 2.9e-16;
Matches 266; Conservative 0; Mismatches 278; Indels 0; Gaps 0;

QY 649 ACTTCCCGGTTTGTGTAAGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 708
DB 652 ACATTTCTTGACATGCTGCGACGATCTGCTATCTGCACTGCGGCTGAGCTTC 593
QY 709 ACCTTGCG 768
DB 592 ACATTCGCGCTCCCAACCACTTTCACCTGCTATGCGCATGCGCGCGCTCACTTATTC 533
QY 769 GAGAGCGGTAAACCAACCG 828
DB 532 GACG 473
QY 829 TCTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 888
DB 472 ATGCCAGCGGTATTTTCAACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 413
QY 889 CGAGACAGTACCTGCG 948
DB 412 GGTGACGAGCGTGTGCTCCCGCGCTTCTTCAAGGAGTGAAGTTTGTGCTTCAAGCGG 353
QY 949 GCTGGGTGCGCAAGGATCTGGGATCTGTTGATCCGCGCTGCTGCTGCTGCTGCTGCTG 1008
DB 352 TCAAGCTGCGCGAAGCTTGTGAACGCGCTGTAATTTGCTGCTGCTGCTGCTGCTGCT 293
QY 1009 GAGCGCATTCGATGATGCGCGGTCTGCGCATGACGAGACTCTCTCTGCTGCTGCTT 1068
DB 292 CGCGCGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 233
QY 1069 ACAACCGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1128
DB 232 TGCATTTCTTCCCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 173
QY 1129 AAGCTGTTCCGATGATGAGGAATGGAAGGCGTTTCCATGCTGCGACGCTGATGAGC 1188
DB 172 AAGCTGCTACCTCTGCTGCGCAAGCTGATGATGATGATGATGATGATGATGATGATG 113
QY 1189 GGCT 1192

Db	112	GTT	109
RESULT 5			
CNS0175Y/c			
LOCUS			
DEFINITION	CNS0175Y	1101 bp	DNA
	Drosophila melanogaster genome survey sequence SP6 end of BAC		
	BACN37106 of DrosBAC library from Drosophila melanogaster (fruit		
	fly), genomic survey sequence.		
ACCESSION	AL108460		
VERSION	AL108460.1		
KEYWORDS	GI:5628764		
SOURCE	GSS.		
ORGANISM	Drosophila melanogaster (fruit fly)		
	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
	Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr		
COMMENT	- Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CPFH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.		
FEATURES			
source	Location/Qualifiers		
	1..1101		
	/organism="Drosophila melanogaster"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:7227"		
	/clone="BACN37106"		
	/clone_11b="DrosBAC"		
	/plasmid="pBelobAC11"		
	/note="end : SP6"		
ORIGIN			
Query Match	2.7%: Score 47.4: DB 9: Length 1101:		
Best Local Similarity	19.1%: Pred. No.0.11,		
Matches	72: Conservative 142; Mismatches 162; Indels 0; Gaps 0;		
QY	793 GGGTTCGCCGAGACGCTTCGCAACTTGAGCGAATCTCTCCACATCGCTCACTGTG	852	
Db	986 BSSBSBSBSSTSSBTSBTSBSSBSSTSSABBSBSBSBSSTSSBTSBTSBTS	927	
QY	853 CCGAAGCGCTGGAGCAATTAGTGGTCCCTTGACGACAGACAGTACCTCGCGCAACG	912	
Db	926 SSSGSSSSBSBSBSBTSBTSBTSBTSBTSBTSBTSBTSBTSBTSBTSBTSBTS	867	
QY	913 TTCTTCGCTCCATGAAAGCTGTTCTTCTTCGCGCGCGCTGGATTGCCAAGGATCG	972	
Db	866 STTTSBSBTSBSYSSBSBTSBTSBTSBTSBTSBTSBTSBTSBTSBTSBTSBTS	807	
QY	973 GATCGTTTGACCGGGTCTGTAACAGCACTGTGTGAGCGCATTCGATATGCGGGT	1032	
Db	806 GSTGSSSCSSBSGKSTSSGSSSTSGTTCCBGSSTGSCCCCTCCTCTGTGTCCGS	747	
QY	1033 CTGGGCAATGAGGAGACGCTCTCTTCGCGCACTTTTACCGGACGCGCTGTCAATGCT	1092	
Db	746 CAGTSSBSKCTYSTSSBSASTSSGSSSTSSGCGTMTGCGCGATTGCTGCTG	687	
QY	1093 GATTACATTGGAGCTGCGACCGCTGCGTGCAGAGTCAAGCTGTTCCGATGGAGAA	1152	
Db	686 CCYBCMTMMCCGSSCCSGCYCGSSKCTSMGTTGCMYCCGCTSCCOTGCGGMAA	627	
QY	1153 TTGGAAGGCGCTTCC	1168	
Db	626 CGTGTGSGCGNKNVM	611	

[illegible]

FEATURES	Source	Location/Qualifiers
ORGANISM	Oncoerhynchus mykiss	
MOL_TYPE	"mrna"	
DB_XREF	"taxon:8022"	
CLONE	"rcay0024b.n.02"	
TISSUE	"adipose tissue, blood, brain, differentiating gonads, gills, interrenal, intestine, kidney, liver, muscle, ovary, pituitary, testis"	
DEV_STAGE	"from embryos to adults"	
LAB_HOST	"DH108"	
CLONE_ID	"ccay"	
NOTES	"Vector: pT73D-Pac; Rainbow trout multi-tissues - normalized + 1 subtraction (rcay) ; Clone distribution : AGENAS Resource centre. Francois PIUMI, Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex, France, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73"	
ORIGIN		
Query Match	2.4%; Score 42.6; DB 5; Length 585;	
Beet Local Similarity	51.9%; Pred. No. 2;	
Matches	96; Conservative 0; Mismatches 89; Indels 0; Gaps 0;	
QY	1085 CGATGGCTGGTTACATTATGGGCTGCACGCGCTGCTCGAGGTCAAGCTGTCGGGTG	1144
DB	165 CTATACCTTTGTAACATTGTGAACCTGTTGATGTGAAGAGATAACTACTTCTTCCA	224
QY	1145 ATGGAAATTGGAAGGCGTTTCCATGCTCCGACGTCATGAGCGCTACTGCGTCTC	1204
DB	225 ATGGTGAAGGCGAGGTCTGATCAAAAGTAAAAACGTTCATCGATATCTGAAGACC	284
QY	1205 CTGAACAATAATGCCAAGCGTTGACGAGAGGCTATTACTGTCGGGTGATGCATCA	1264
DB	285 CTGGAAGACAGCAGAGCGCCCTGGACAAAGTGGCTGTACATACACCGAGACATTGGCA	344
QY	1265 AATGG 1269	
DB	345 AATGG 349	
RESULT 9		
LOCUS	BX882027	
DEFINITION	795 bp mRNA linear EST 20-JUN-2004	
ACCESSION	BX882027 tcdh Oncoerhynchus mykiss cDNA clone tcdh0053c.k.18 spim,	
VERSION	1	
KEYWORDS	EST.	
SOURCE	Oncoerhynchus mykiss (rainbow trout)	
ORGANISM	Oncoerhynchus mykiss	
REFERENCE	1 Gotoyoun,M., Guiguen,Y. and Le Gac,F. Construction and primary characterization of normalized cDNA libraries in rainbow trout, Oncoerhynchus mykiss Unpublished (2003)	
AUTHORS	On Dec 18, 2003 this sequence version replaced gi:40126280.	
TITLE	Contact: Guiguen Y	
JOURNAL	INRA - SCRIBS	
COMMENT	Campus de Beaulieu, RENNES cedex, 35042, France Tel: 02.23.48.50.09 Fax: 02.23.48.50.20 Email: Yann.Guiguen@beaulieu.rennes.inra.fr Sequence cleaned of vector, adaptor and repetitions. Contact us at sigenasupport@jouy.inra.fr to obtain the chromatogram of this sequence. Plate:0053 row: k column: 18	

```

FEATURES
source
Seq primer: M13r.
location/Qualifiers
1..795
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcdbK053c.k.18"
/tissue_type="multi-tissues"
/dev_stage="from embryos to adults"
/lab_host="DH10B"
/clone_lib="tcdbk"
/note="vector: pRTT3D-pac; AGENAB Rainbow trout
multi-tissues - normalized + 2 subtractions; Clone
distribution: AGENAB Resource centre, Francois Piumi,
Francois.Piumi@joy.irsra.fr, IRSRA, CEA radiobiologie et
Etude du genome (IREG), Domaine de Vilvert, 78382,
Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
(0) 1.34.65.22.73"

```

	Query Match	2.4%	Score 42.6;	DB 5;	Length 795;
	Best Local Similarity	51.9%;	Pred. No. 2.1;		
	Matches	96;	Conservative	0;	Mismatches 89; Indels 0; Gaps 0;
QY	1085	CGATGCGCTGTTAACAATTGGAGCTGCACGAGCCCTGCGCTGGAGATCAAGCTGTTCCGAGTC			1144
Db	72	CTATACCTCTTAAACATTGTGAACCTGGTTAGTTGAAGAGATGAACCTTGTGCTTCA			131
QY	1145	ATGGGAAATTGGAAAGGCGCTTTCATGCTCCGCACGTATGAGCGGCTACTCGGCTGCT			1204
Db	132	ATGGTGAAGGGAGAGCTGTATCAAAAGGTAAACAGTCTTCATCGAATATCTGAAAGACC			191
QY	1305	CTGAACAATAATGCCACAGCTTTCGACGAGAGAGCTATTACTGCTCCGGTATGTCATCA			1264
Db	192	CTGAGAGACACGACAGGCGCTCGACAAATAGTGTGCTACACACCGAGACATTGGCA			251
QY	1265	AATTGG 1269			
Db	252	AATGG 256			

RESULT	10
LOCUS	BX911066
DEFINITION	BX911066 tcdk Oncorhynchus mykiss cDNA clone tcdk0066c.d.13 5prim,
ACCESSION	BX911066
VERSION	BX911066
KEYWORDS	BX911066.2 GI:43416323
SOURCE	EST.
ORGANISM	Oncorhynchus mykiss (rainbow trout)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus. 1 (bases 1 to 815) Govoroun,M., Guiguen,Y. and Le Gac,F. Construction and primary characterization of normalized cDNA libraries in rainbow trout, Oncorhynchus mykiss unpublished (2003) On Jan 22, 2004 this sequence version replaced gi:41127874.
JOURNAL COMMENT	Contact: Guiguen Y INRA - SCRIBS Campus de Beaulieu, RENNES cedex, 35042, France Tel.: 02.23.48.50.09 Fax: 02.23.48.50.20 Email: Yann.Guiguen@beaulieu.rennes.inra.fr Sequence cleaned of vector, adaptor and repetitions. Contact us at sigenasupport@jouy.inra.fr to obtain the chromatogram of this sequence Plate: 0068 row: d column: 13 Seq primer: M13R. Location/Qualifiers 1..815
FEATURES	
SOURCE	

```

/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcbk0068c.d.13"
/tissue_type="multi-tissues"
/developmental_stage="from embryos to adults"
/lab_host="DH10B"
/clone_id="tcbk"
/notes="vector: pT73D-pac; AGENAE Rainbow trout
multi-tissues - normalized + 2 subtractions; Clone
distribution : AGENAE Resource centre, Francois Piumi,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (LRE6), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
(0) 1.34.65.22.73"

```

Query Match	Score	DB	Length
Best Local Similarity	51.9%	Pred. No. 2.2	
Matches	96	Conservative	0
		Mismatches	89
		Indels	0
		Gaps	0
Qy	1085	CGATGCGTGTTCATCTTGGCGCTGCAGCGCTGCAGCTGCAGCTCAAGCTGTTCCGCTG	1144
Db	359	CTATACCTTGTAAACATGTGTAAACCTGTTGATGTTGAAGAGATGAACACTTTCCTTCCA	418
Qy	1145	ATGGCAATTGCAAGGGCGTTCCATGTCCTCCGACGTCATGAGGGCGCTACTGGCGTCTC	1204
Db	419	ATGTGTAAAGGGGAGGCTCTGTATCAAAAGGTAAACGTTTATCGATATCTGAAAGACC	478
Qy	1205	CTGAACAAATGCCCAAGCGTTTCGACGAGGAGCTATTACTGCTCCGCGTATGCATCA	1264
Db	479	CTGAGAAGACAGCAGAGGCGCTTGACAAAGATGGCTGCTACACACCGAGACATTGSCA	538
Qy	1265	AATTG	1269
Db	539	AATGG	543

RESULT 11	
LOCUS	CNS0004NB 839 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TIR3 end of BAC # BACR016 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL054280
VERSION	AL054280.1 GI:4931788
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peiryocta; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 839) Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefegenoscope.cns.fr
JOURNAL	- Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library was prepared by Kazutoyo Osoegawa and Aaron Mammone in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .

FEATURES	source	location/Qualifiers
		1..839
		/organism="Drosophila melanogaster"
		/mol_type="genomic DNA"
		/db_xref="taxon:7227"
		/clone="BACR10E16"
		/clone_1fb="RPC1-98"
		/note="Tend : TET3"
ORIGIN		
Query Match	2.3%; Score 41.4; DB 9; Length 839;	
Best Local Similarity	18.4%; Pred.No.4.7;	
Matches	69; Conservative 129; Mismatches 176; Indels 0; Gaps 0;	
OY	746 ACAACGGCGGACACTACTGACGACGCGTAACCAACCGCCCAAGGTTCCCGCAGA	805
Db	430 ANNACCCACMCCNNATATNTNTTAAAGAAATATSMSSSACSAASGCSASCVAA	489
OY	806 CGCTTCGCACTTGAGCGAAATCTCTCCCACTGCTACTCTAGTGGCCGAAGGCTGG	865
Db	490 CSSACAGSAGSSSSAGSSSAGSCGAGSSSAGSGAGSGRSGSGAGSGSGGSGRSG	549
OY	866 AGGATTTAGTGGTGCCCTTCGACGACAGACGATACCTCGCGAACGCTTTCTTCGCGCA	925
Db	550 CCGAVGASAGSSSSAGCSASSASSSSSAGSCMSGSSSSSSSMASSCSSSSSGTAATCS	609
OY	926 TGAAGCTGTTCTTCTTCGCGCGCGCTGGTGTGTCGAAGGATCTGGATCGTTGAGCC	985
Db	610 GSGGAGVGSAAAVASRGARGGGAGAGSRAVSAASAAASASASAVGAACSGGS	669
OY	986 GCGTTCGCTGAACACGACTGTGTGTAAGCCCATTTCCGATATGCGGCTGTGGCATGACG	1045
Db	670 GASSSSSGAARARCAVCASSASSGSCGSSSVCASVSSGAASGASGCGCGSSSG	729
OY	1046 AGACTGCTCCTTCCTCGACATTTTACACCGACCGCTGCTGATGAGCTGTTACATTGGCC	1105
Db	730 CSGSCSSSSSSAGSSSSGCGSSGCGSSCGSSCGSSCGSSCGSSCGSSCGSSCGSS	789
OY	1106 TGCCAGCGCGCTGCG 1119	
Db	790 SSSCSVSCSMGSC 803	
RESULT 12		
LOCUS	BM442982	
DEFINITION	EB-ro02_SB002_B12.R rooc. 3 week, hydroponic grown, low nitrogen, cv	
ACCESSION	BM442982	
VERSION	BM442982.2	
KEYWORDS	EST.	
SOURCE	Hordeum vulgare subsp. vulgare	
ORGANISM	Hordeum vulgare subsp. vulgare	
REFERENCE	Bukariyotai, Vitridiplantae; Striptophyta; Embryophyta; Tracheophyta;	
AUTHORS	Healey, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,	
TITLE	Ramsey, L., Machray, G., Marshall, D.F.M. and Waugh, R.	
JOURNAL	Development of barley transcriptome Resources	
COMMENT	Unpublished (2001)	
	On Feb 1, 2002 this sequence version replaced gi:18473745.	
	Contact: Waugh R, Marshall DF	
	Genome Dynamics/Computational Biology	
	Scottish Crop Research Institute	
	Invergowrie, Dundee, DD2 5DA, Scotland, UK	
	Tel: 00 44 1382 562731	
	Fax: 00 44 1382 562426	
	Email: est@scri.sari.ac.uk	
	All sequence has a phred quality score of 20 or over	
	Seq primer: M13 reverse	
	Location/Qualifiers	

	source	1..240	/organism="Hordeum vulgare subsp. vulgare"
		/mol_type="mRNA"	
		/culivar="Optic"	
		/sub_species="vulgare"	
		/db_xref="taxon:112509"	
		/clone="EBR002_SQ002_B12"	
		/tissue_type="root"	
		/dev_stage="3 week"	
		/lab_host="DH10B"	
		/clone_id="root, 3 week, hydropenic grown, low nitrogen, cv Optic, EBR002"	
		/note="Vector: pSPORT1; Site 1: Sal I; Site 2: Not I; Non-normalised library, directionally cloned into pSPORT1. Derived from roots of 3 week old Nitrogen stressed barley plants. Developed as part of the barley transcriptome resources of BBSRC/SERAD funded cereal IGF (Investigating Gene Function) project."	
ORIGIN			
Query Match	2.3%	Score 41.2;	DB 4; Length 240;
Best Local Similarity	56.7%;	Pred. No. 3.9;	
Matches	76; Conservative	0; Mismatches	58; Indels 0; Gaps 0;
Oy	551	TTGGCAAGTCTTGTCACCTTGCGCTGACCAAACTGCCTAAGCGGTGCCACTACTC	610
Db	39	TGGCCCTGTTCTGCACACCCTCCGGACAGACCCGCCAACAGGTGTGCCGTGACGC	98
Oy	611	AGCGAATGCTCTGCGCCCAATCAGACAGATGCTTCTGCAAACTTTCCCGATTTTTGTAAG	670
Db	99	AGCGGAACCTGCGCGCTCGGTCCAGAACATCCGGCGGTGTACCGGTTGTGACAGCG	158
Oy	671	AGCCGCGGTGCTG 684	
Db	159	ACGCGACGGTGTG 172	
RESULT 13			
EM443022			
LOCUS			
DEFINITION	BM443022	363 bp	mRNA linear EST 23-JUN-2002
	EBR002 SQ002 D13 R root, 3 week, hydropenic grown, low nitrogen, cv		
	Optic_EBR002 Hordeum vulgare subsp. vulgare cDNA clone		
	EBR002 SQ002_D13 5', mRNA sequence.		
ACCESSION	BM443022		
VERSION	BM443022.2	GI:21947674	
KEYWORDS	EST.		
SOURCE	Hordeum vulgare subsp. vulgare		
ORGANISM	Hordeum vulgare subsp. vulgare		
	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
	Pooidae; Triticeae; Hordeum.		
	1 (bases 1 to 363)		
AUTHORS	Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,		
REFERENCE	Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.		
JOURNAL	Development of Barley Transcriptome Resources		
COMMENT	Unpublished (2001)		
	On Feb 1, 2002 this sequence version replaced gi:18473785.		
	Contact: Waugh R, Marshall DF		
	Genome Dynamics/Computational Biology		
	Scottish Crop Research Institute		
	Invergowrie, Dundee, DD2 5DA, Scotland, UK		
	Tel: 00 44 1382 562731		
	Fax: 00 44 1382 562426		
	Email: est@scri.sari.ac.uk		
	All sequence has a Phred quality score of 20 or over		
	Seq primer: M13 reverse.		
FEATURES			
source	Location/Qualifiers		
	1..363		
	/organism="Hordeum vulgare subsp. vulgare"		
	/mol_type="mRNA"		
	/culivar="Optic"		
	/sub_species="vulgare"		
	/db_xref="taxon:112509"		

/clone="EBR002_S0002_D13"
/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH103"
/clone.lib="root", 3 week, hydroponic grown, low nitrogen,
cv Optic, EBR002"
/note="Vector: pSPORT1; Site 1: Sal I; Site 2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from roots of 3 week old Nitrogen stressed barley
plants. Developed as part of the barley transcriptome
resources of BBSRC/SEERAD funded cereal IGR (Investigating
Gene Function) Project."

ORIGIN

Query Match 2.3%; Score 41.2; DB 4; Length 363;
Best Local Similarity 56.7%; Pred. No. 4.3;
Matches 76; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 551 TTGCCAAGTTCTTGTTCATTCTTGGCTCTTACCAAACTGCTTAAAGCGGTCGCACTACTC 610
DB 39 TGCCCTGTTCTGCAACCTCCGCGACGAGCCGCAAGGAGTGTCCGCTGACGC 98
QY 611 AGGAATGCTCTGGCGCAATCAGAGATGCTTCTGCAACTTCCCGTTTGTGTAAG 670
DB 99 AGGCAACCTCGCGCCTCCGTCAGAACATCCGCGGCTGTACCGGTTGTGAGACGG 158
QY 671 AGCGCGCGTGTG 684
DB 159 ACGCGACGCTGTG 172

RESULT 14
LOCUS CR676208 856 bp mRNA linear HTC 12-AUG-2004
DEFINITION Tetraodon nigroviridis full-length cDNA.
ACCESSION CR676208.1 GI:51173964
VERSION CR676208.1
KEYWORDS HTC; cDNA; full-length; Tetraodon nigroviridis.
SOURCE Tetraodon
ORGANISM Tetraodon

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae.
1 (bases 1 to 856)
Genoscope.
Direct Submission

JOURNAL Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -

COMMENT (E-mail : segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
The sequences are based on single pass reads.
More information available at
http://www.genoscope.cns.fr/tetraodon.

FEATURES

source

1..856
/organism="Tetraodon"
/mol_type="mRNA"
/db_xref="taxon:47144"
/tissue_type="Muscle"

ORIGIN

Query Match 2.3%; Score 40.8; DB 3; Length 856;
Best Local Similarity 56.8%; Pred. No. 7;
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 448 GCGGAATTGCGCGCGCGCGAGGTGATTTGACAGCCTGCTGAGAGCCTGTGGG 507
DB 657 GTCAAGATGCTCAAGAGCCCAAGTTGAGTGGGCAACTGATGAGCTCCACGGTGA 716
QY 508 ATTGAGCAGATATGCTTGTGGCGCAACTGCCCGCATACGATTTGCCAAGTTCTTGTTC 567
DB 717 GGTGGGCGAGAACTGCTCTAAGGCACTGTGACGATTACTGAGCCCAAGTGAAGCGG 776

QY 568 ACTTCTGGCTCT 579
DB 777 GCTGATGGCTAT 788

RESULT 15
LOCUS CR662673 873 bp mRNA linear HTC 11-AUG-2004
DEFINITION Tetraodon nigroviridis full-length cDNA.
ACCESSION CR662673
VERSION CR662673.1 GI:51159118
KEYWORDS HTC; cDNA; full-length; Tetraodon nigroviridis.
SOURCE Tetraodon
ORGANISM Tetraodon

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae.
1 (bases 1 to 873)
Genoscope.
Direct Submission

JOURNAL Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -

COMMENT (E-mail : segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
The sequences are based on single pass reads.
More information available at
http://www.genoscope.cns.fr/tetraodon.

FEATURES

source

1..873
/organism="Tetraodon"
/mol_type="mRNA"
/db_xref="taxon:47144"
/tissue_type="Muscle"

ORIGIN

Query Match 2.3%; Score 40.8; DB 3; Length 873;
Best Local Similarity 56.8%; Pred. No. 7;
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 448 GCGGAATTGCGCGCGCGCGAGGTGATTTGACAGCCTGCTGAGAGCCTGTGGG 507
DB 666 GTCAAGATGCTCAAGAGCCCAAGTTGAGTGGGCAACTGATGAGCTCCACGGTGA 725
QY 508 ATTGAGCAGATATGCTTGTGGCGCACTGCCCGCATACGATTTGCCAAGTTCTTGTTC 567
DB 726 GGTGGGCGAGGCACTGCTCTAAGGCACTGTGACGATTACTGAGCCCAAGTGAAGCGG 785
QY 568 ACTTCTGGCTCT 579
DB 786 GCTGATGGCTAT 797

Search completed: February 7, 2005, 21:35:19
Job time : 5109 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2005, 20:09:50 ; Search time 979 Seconds
(Without alignments)
10388.378 Million cell updates/sec

Title: US-09-750-986D-29

Perfect score: 1770
Sequence: 1 atgcgtctctcgcagcgcgc.....gtcagcagcgcacactgtga 1770

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCRUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1770	100.0	1770	8	US-08-976-063C-29
2	1770	100.0	1770	8	US-09-750-986D-29
3	1770	100.0	32679	8	US-08-976-063C-1
4	1770	100.0	32679	10	US-09-750-986D-1
5	527.8	29.8	1725	15	US-10-369-493-44177
6	398.4	22.5	1728	15	US-10-369-493-35384
7	398.4	22.5	1728	15	US-10-369-493-38427
8	398.4	22.5	1728	15	US-10-369-493-38903
9	372	21.0	1660	15	US-10-369-493-38007
10	225.6	12.7	1866	15	US-10-369-493-40734
11	57.4	3.2	1879	18	US-10-716-803-5
12	46.8	2.6	536	16	US-10-338-110-119

13	46.8	2.6	1566	15	US-10-369-493-42307	Sequence 42307, A
14	46.4	2.6	1644	16	US-10-282-122A-30430	Sequence 30430, A
15	45.4	2.6	6522	16	US-10-282-122A-31533	Sequence 31533, A
16	45.2	2.6	1977	16	US-10-369-493-37463	Sequence 37463, A
17	44.4	2.5	1632	16	US-10-282-122A-40967	Sequence 40967, A
18	43.4	2.5	342	16	US-10-282-122A-1896	Sequence 1896, A
19	43.4	2.5	410	16	US-10-282-122A-3010	Sequence 3010, A
20	43.4	2.5	1701	15	US-10-369-493-24535	Sequence 24535, A
21	43.4	2.5	1701	16	US-10-282-122A-6822	Sequence 6822, A
22	42.4	2.4	1677	15	US-10-369-493-41695	Sequence 41695, A
23	42	2.4	1701	15	US-10-369-493-41079	Sequence 41079, A
24	41.6	2.4	825	17	US-10-437-963-73949	Sequence 73949, A
25	41.6	2.4	882	14	US-10-184-644-574	Sequence 574, A
26	41.6	2.4	882	14	US-10-184-644-574	Sequence 574, A
27	41.4	2.3	1680	15	US-10-369-493-33018	Sequence 33018, A
28	41.4	2.3	1680	15	US-10-369-493-33087	Sequence 33087, A
29	41.4	2.3	1803	17	US-10-437-963-43352	Sequence 43352, A
30	40.6	2.3	1500	15	US-10-369-493-43198	Sequence 43198, A
31	40.6	2.3	13029	9	US-09-815-242-4052	Sequence 4052, A
32	40.6	2.3	13029	16	US-10-282-122A-7240	Sequence 7240, A
33	40.4	2.3	1049	14	US-10-123-155-358	Sequence 358, A
34	40.4	2.3	1049	15	US-10-146-731-358	Sequence 358, A
35	40.4	2.3	1049	15	US-10-140-472-358	Sequence 358, A
36	40.4	2.3	1049	15	US-10-141-761-358	Sequence 358, A
37	40.4	2.3	1049	15	US-10-142-885-358	Sequence 358, A
38	40.4	2.3	1049	15	US-10-158-790-358	Sequence 358, A
39	40.4	2.3	1049	15	US-10-137-871-358	Sequence 358, A
40	40.4	2.3	1049	15	US-10-140-923-358	Sequence 358, A
41	40.4	2.3	1049	15	US-10-141-756-358	Sequence 358, A
42	40.4	2.3	1049	15	US-10-141-759-358	Sequence 358, A
43	40.4	2.3	1049	15	US-10-140-805-358	Sequence 358, A
44	40.4	2.3	1049	15	US-10-140-864-358	Sequence 358, A
45	40.4	2.3	1049	16	US-10-142-426-358	Sequence 358, A

ALIGNMENTS

RESULT 1
US-08-976-063C-29
; Sequence 29, Application US/08976063C
; Publication No. US20020182697A1
GENERAL INFORMATION:
APPLICANT: Alexander Steinhuechel; Horst Priefert; Jürgen Rabenhorst
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
TITLE OF INVENTION: CONFERYL ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN A
TITLE OF INVENTION: ACID AND THEIR USE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SPRING KRAMER SCHAEFER & BRISCOE
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: HP VECTRA
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976, 063C
FILING DATE: 21-NOV-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 196 49 655.1 (Germany)
FILING DATE: 29-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700

```

? TELEFAX: (914) 332-1844
?
? TELEX:
? INFORMATION FOR SEQ ID NO: 29:
?
? SEQUENCE CHARACTERISTICS:
?     LENGTH: 1770 base pairs
?     TYPE: nucleic acid
?     STRANDEDNESS: double
?     TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEetical: NO
? ANTI-SENSE: NO
? FEATURE:
?     NAME/KEY: CDS
?     LOCATION: 1..167
? OTHER INFORMATION: /product=
? OTHER INFORMATION: "Ferrihaemre-coA-Synthetase"
? OTHER INFORMATION: /gene="fcs"
?
? US-08-976-063C-29

```

Query Match	100.0%;	Score 1770;	DB 8;	length 1770;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1770;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	ATGGTTCCTCGAAGCGCTTCCTCCCTCCGGGTGCAATTCTTGACGCTCGACAT	60
Db	1	ATGGTTCCTCGAAGCGCTTCCTCCCTCCGGGTGCAATTCTTGACGCTCGACAT	60
Qy	61	TGGGCTAAGACCCGCTCCAGAACAACTGCGTGTGCTGCAGGCGGCAATGGGAAATGG	120
Db	61	TGGGCTAAGACCCGCTCCAGAACAACTGCGTGTGCTGCAGGCGGCAATGGGAAATGG	120
Qy	121	CGTGATACGCTACGGGGAATGTGCCAACGTCGGCGGCATGGCAGAGCTTGCTT	180
Db	121	CGTGATACGCTACGGGGAATGTGCCAACGTCGGCGGCATGGCAGAGCTTGCTT	180
Qy	181	CCTTACGGACTATCGGAGAGCGTCGCTGCTTATCGTCTGTGAAATAGCCTGGAACAT	240
Db	181	CCTTACGGACTATCGGAGAGCGTCGCTGCTTATCGTCTGTGAAATAGCCTGGAACAT	240
Qy	241	CTTCAGCTGGGATTTGGGGGCTATGATATGCGGGACATTCCTTATTGGCCGCTGTCTCTCT	300
Db	241	CTTCAGCTGGGATTTGGGGGCTATGATATGCGGGACATTCCTTATTGGCCGCTGTCTCTCT	300
Qy	301	TATTCACCTGCTGTGCAAGATTTTGGCGAAGCTGACATCGTATGCTTCTTGCAACCG	360
Db	301	TATTCACCTGCTGTGCAAGATTTTGGCGAAGCTGACATCGTATGCTTCTTGCAACCG	360
Qy	361	GGACTGCTCTTGTGCTGCCGATGCAACCTTTCCAGCGCGCAATTGAGACCATTCGCGC	420
Db	361	GGACTGCTCTTGTGCTGCCGATGCAACCTTTCCAGCGCGCAATTGAGACCATTCGCGC	420
Qy	421	GAGCAGCTGCCCGCAATCTTCACCTCGAGGGCAATTGGCGCGGGCGGACGCTGAGTTT	480
Db	421	GAGCAGCTGCCCGCAATCTTCACCTCGAGGGCAATTGGCGCGGGCGGACGCTGAGTTT	480
Qy	481	GACAGCCTGTGAGCAGCCTGTGTGGGATTTAGGACAGATTAATGCTTTGCGGCACTGGC	540
Db	481	GACAGCCTGTGAGCAGCCTGTGTGGGATTTAGGACAGATTAATGCTTTGCGGCACTGGC	540
Qy	541	CCCGATACGATTTGCCAAGTTCTTGTTCACTTCTGGCTTACCAACCTGCTTAAAGCGGTG	600
Db	541	CCCGATACGATTTGCCAAGTTCTTGTTCACTTCTGGCTTACCAACCTGCTTAAAGCGGTG	600
Qy	601	CCGACTACTACGGGAATGCTCTGGGCGCAATAGCAGATGCTTTGCGCAAACTTCCGGT	660
Db	601	CCGACTACTACGGGAATGCTCTGGGCGCAATAGCAGATGCTTTGCGCAAACTTCCGGT	660
Qy	661	TTTGGTGAAGACCGCGGCTGCTGTGAGCTGTGTGCGTGGAAACAACCTTCCGCGGC	720
Db	661	TTTGGTGAAGACCGCGGCTGCTGTGAGCTGTGTGCGTGGAAACAACCTTCCGCGGC	720
Qy	721	AGCCACAACATCGGACTGTGTGTATACAGGGCGGACGTACTTACCTTGACGACGCTTAA	780
Db	721	AGCCACAACATCGGACTGTGTGTATACAGGGCGGACGTACTTACCTTGACGACGCTTAA	780

Db	721	AGCCACAACATCGGCATCGTGTGACCAAGCGGCGACAGTACTA	CTTGAACGACGGTAA	780
Qy	781	CCAAACGCCCAAGGGTTCCGCGAGACGCTTCGCACTTGAAGCAAA	ACTCTCCACATGCG	840
Db	781	CCAAACGCCCAAGGGTTCCGCGAGACGCTTCGCACTTGAAGCAAA	CTCTCCACATGCG	840
Qy	841	TACCTCACTGTGCGAAGGCTGGAGAGATTAGTGGGTCCCTTGA	CCAGACAGTATAC	900
Db	841	TACCTCACTGTGCGAAGGCTGGAGAGATTAGTGGGTCCCTTGA	CCAGACAGTATAC	900
Qy	901	CTGCGGCAACGCTTCTTCGCTCGCATGACGTGTTCTTTCGCG	CGGCTGGGTGTG	960
Db	901	CTGCGGCAACGCTTCTTCGCTCGCATGACGTGTTCTTTCGCG	CGGCTGGGTGTG	960
Qy	961	CAAGGATCTGGGATCGTTTGAACGGGTGCTGAACAGACTGTG	TGAGCGCATTCG	1020
Db	961	CAAGGATCTGGGATCGTTTGAACGGGTGCTGAACAGACTGTG	TGAGCGCATTCG	1020
Qy	1021	ATGATGCGGGCTCTGGCATGACGAGACTCTCTCTCTGCACTT	TTACACCGACG	1080
Db	1021	ATGATGCGGGCTCTGGCATGACGAGACTCTCTCTCTGCACTT	TTACACCGACG	1080
Qy	1081	CTGTCCATGGCTGTGTTACATTTGGGCTGCCAGCGCTTGGCT	GGAGGTCAAGCTGTTCCG	1140
Db	1081	CTGTCCATGGCTGTGTTACATTTGGGCTGCCAGCGCTTGGCT	GGAGGTCAAGCTGTTCCG	1140
Qy	1141	GTCGATGGAAATTTGGAAGGCGGTTTCATGATCCGCACTCA	TGACCGGCTACTGGCGT	1200
Db	1141	GTCGATGGAAATTTGGAAGGCGGTTTCATGATCCGCACTCA	TGACCGGCTACTGGCGT	1200
Qy	1201	GCTCTGAAACAAATATGCCAAGCGGTTCGACGAGGAAGGCT	ATTACTGCTCCGGTATGCC	1260
Db	1201	GCTCTGAAACAAATATGCCAAGCGGTTCGACGAGGAAGGCT	ATTACTGCTCCGGTATGCC	1260
Qy	1261	ATCAAAATTTGGCAATCCGCGCATCTCGAAGAGTCTGATGTT	TGAACGGTCCAAATTGCT	1320
Db	1261	ATCAAAATTTGGCAATCCGCGCATCTCGAAGAGTCTGATGTT	TGAACGGTCCAAATTGCT	1320
Qy	1321	GAAGACTTCAAGCTGTCTCAGGGGTATTTGTACAGCTTGG	CCATTGGCCGACGCGGCG	1380
Db	1321	GAAGACTTCAAGCTGTCTCAGGGGTATTTGTACAGCTTGG	CCATTGGCCGACGCGGCG	1380
Qy	1381	GTTCTGGAAGCGGCTCTTAACGTCCTGACGTAAGTGTGTC	GCTGCTGATCTGAAATGC	1440
Db	1381	GTTCTGGAAGCGGCTCTTAACGTCCTGACGTAAGTGTGTC	GCTGCTGATCTGAAATGC	1440
Qy	1441	CTTGAATGCTCGTGTTCCGGGCTTCTGACCTGCGCGTAC	CTGTCGGGGCTTAGGAAA	1500
Db	1441	CTTGAATGCTCGTGTTCCGGGCTTCTGACCTGCGCGTAC	CTGTCGGGGCTTAGGAAA	1500
Qy	1501	GAGGCGTCCGACCGCAGGTGTCTTGCACATGAGCCGAT	TTCGAGCTGACATGCG	1560
Db	1501	GAGGCGTCCGACCGCAGGTGTCTTGCACATGAGCCGAT	TTCGAGCTGACATGCG	1560
Qy	1561	CTCAAAAGCATCAATGAGAAAGCACTGGATATGCAATG	TGAGGCTTCAAGGCTC	1620
Db	1561	CTCAAAAGCATCAATGAGAAAGCACTGGATATGCAATG	TGAGGCTTCAAGGCTC	1620
Qy	1621	CTCGATACGCGCGCGTCCGATTTGAAGGGGAGAGTCACT	GACAAAGGCTCGATCAAC	1680
Db	1621	CTCGATACGCGCGCGTCCGATTTGAAGGGGAGAGTCACT	GACAAAGGCTCGATCAAC	1680
Qy	1681	CGCGCTGTTTTCGACATGCGCGTCCGCGAAAGTTGATG	CGCTGTATCGTGTGAAGTCAA	1740
Db	1681	CGCGCTGTTTTCGACATGCGCGTCCGCGAAAGTTGATG	CGCTGTATCGTGTGAAGTCAA	1740
Qy	1741	TCCATGCTGTGTACGACAGGCCCAACTGTGA	1770	
Db	1741	TCCATGCTGTGTACGACAGGCCCAACTGTGA	1770	

RESULT 2
US-09-750-986D-29

Sequence 29. Application US/09750986D
Publication No. US20030228670A1
GENERAL INFORMATION:
APPLICANT: Steindubchel, Alexander
APPLICANT: Priefert, Horst
APPLICANT: Rabenhorst, Jurgen
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
FILE REFERENCE: Bayer-9998-CAO
CURRENT APPLICATION NUMBER: US/09/750, 986D
PRIORITY FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 29
LENGTH: 1770
TYPE: DNA
ORGANISM: Pseudomonas sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (1..(1767))
OTHER INFORMATION: product = "Ferulasaure-CoA-Synthetase" / gene =
OTHER INFORMATION: "fcgs"
US-09-750-986D-29

Query Match 100.0%; Score 1770; DB 10; Length 1770;
Best local similarity 100.0%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCTCTCGAGGGGCTTCTCCCTCCGGGTGGAATTTCTTGAGCGCTGAGAT 60
DB 1 ATGGCTCTCTCGAGGGGCTTCTCCCTCCGGGTGGAATTTCTTGAGCGCTGAGAT 60
QY 61 TGGGCTAAGACCCCTCCAGAAACAACCTGCTGCTGCGAGGGCGGCAATAGGGAAATG 120
DB 61 TGGGCTAAGACCCCTCCAGAAACAACCTGCTGCTGCGAGGGCGGCAATAGGGAAATG 120
QY 121 CGTCGATACAGCTACGCGGAAATGTTCCACAACGTCGCGGCATCGACAGAGCTGCTT 180
DB 121 CGTCGATACAGCTACGCGGAAATGTTCCACAACGTCGCGGCATCGACAGAGCTGCTT 180
QY 181 CCTTACCGACTATATGCGAGAGCGCTCGTGTATATGCTCTTGAAATGAACTTGAAAT 240
DB 181 CCTTACCGACTATATGCGAGAGCGCTCGTGTATATGCTCTTGAAATGAACTTGAAAT 240
QY 241 CTTTACGCTGCAATTTGGGGCTATATGCGGGCAATCCCTATTTGCCGGTCTCTGCT 300
DB 241 CTTTACGCTGCAATTTGGGGCTATATGCGGGCAATCCCTATTTGCCGGTCTCTGCT 300
QY 301 TATTCAGCTGCTGTCGCAAGATTTGGCGAAGCTGCTGCATCGTAGGTCTTTCTGCAAC 360
DB 301 TATTCAGCTGCTGTCGCAAGATTTGGCGAAGCTGCTGCATCGTAGGTCTTTCTGCAAC 360
QY 361 GAGCTGCTTTGCTGCGATGAGCACTTTCCAGCGCGCAATTTGAGAACCAATTCGCG 420
DB 361 GAGCTGCTTTGCTGCGATGAGCACTTTCCAGCGCGCAATTTGAGAACCAATTCGCG 420
QY 421 GAGCAGCTGCGCGAATCTTCACTCGAGCGCAATTTGGCGGGCGGCGACGCTGAGTTT 480
DB 421 GAGCAGCTGCGCGAATCTTCACTCGAGCGCAATTTGGCGGGCGGCGACGCTGAGTTT 480
QY 481 GACAGCTGCTGAGAGAGCTGCTGAGATTGAGGCAAGTAATGCTTTGCGGCAACTGCG 540
DB 481 GACAGCTGCTGAGAGAGCTGCTGAGATTGAGGCAAGTAATGCTTTGCGGCAACTGCG 540
QY 541 CCGCATAGATTTGCAAGTTTGTTCATCTTGGCTTACCAAACTGCTTAAAGCGGTG 600
DB 541 CCGCATAGATTTGCAAGTTTGTTCATCTTGGCTTACCAAACTGCTTAAAGCGGTG 600
QY 601 CCGACTCTCAGCAATGCTCTGCGCAATCAGAGATGCTTCTGCAAACTTTCCGGTT 660
DB 601 CCGACTCTCAGCAATGCTCTGCGCAATCAGAGATGCTTCTGCAAACTTTCCGGTT 660

DB 601 CCGACTCTCAGCAATGCTCTGCGCAATCAGAGATGCTTCTGCAAACTTTCCGGTT 660
QY 661 TTTGGTGAAGAGCCCGGCTGCTGAGACTGTTGCCGTGGAACCACTTCCGGGC 720
DB 661 TTTGGTGAAGAGCCCGGCTGCTGAGACTGTTGCCGTGGAACCACTTCCGGGC 720
QY 721 AGCCACAACATCGGCATCGTGTGTGAACAAGCGGCGACGTAACCTTGAAGAGAGTAA 780
DB 721 AGCCACAACATCGGCATCGTGTGTGAACAAGCGGCGACGTAACCTTGAAGAGAGTAA 780
QY 781 CCAACCGCCAGAGGTTGCGCGAGACGCTTCGCAACTTGAAGCAAACTCTCCACTGCG 840
DB 781 CCAACCGCCAGAGGTTGCGCGAGACGCTTCGCAACTTGAAGCAAACTCTCCACTGCG 840
QY 841 TACCTCATGTCGCGAAAGGCTGGAAGAAATTAAGTGGTCCCTTGAAGCAAGTAAC 900
DB 841 TACCTCATGTCGCGAAAGGCTGGAAGAAATTAAGTGGTCCCTTGAAGCAAGTAAC 900
QY 901 CTGCGCAACGCTTCTTCCGCTGCAATGAGCTGTTCTTCTGCGGCGGCTGAGTTGCG 960
DB 901 CTGCGCAACGCTTCTTCCGCTGCAATGAGCTGTTCTTCTGCGGCGGCTGAGTTGCG 960
QY 961 CAAGGATCTGGAGTCGTTTGAACCGGCTCGTGAAGACGACTGTGTGAGCGCAATTGCG 1020
DB 961 CAAGGATCTGGAGTCGTTTGAACCGGCTCGTGAAGACGACTGTGTGAGCGCAATTGCG 1020
QY 1021 ATGATGGCGGCTCTGGGCTATGACGAGACGCTCTTCTTCTGCACTTTTACACCGGAC 1080
DB 1021 ATGATGGCGGCTCTGGGCTATGACGAGACGCTCTTCTTCTGCACTTTTACACCGGAC 1080
QY 1081 CTGTCATGAGCTGTATCATTTGGGCTGCAAGCGCTGCTGCGAGTCAAGCTGTTCCG 1140
DB 1081 CTGTCATGAGCTGTATCATTTGGGCTGCAAGCGCTGCTGCGAGTCAAGCTGTTCCG 1140
QY 1141 GTTCATGGAATTTGAAGGCGCTTTCATGTCGACGTCATGACGCGCTACTGCGCT 1200
DB 1141 GTTCATGGAATTTGAAGGCGCTTTCATGTCGACGTCATGACGCGCTACTGCGCT 1200
QY 1201 GCTTCTGAACAAATGCGCCAAAGCTTTCAGAGAGAAAGCTTATCTCTCCGGTGAATGCC 1260
DB 1201 GCTTCTGAACAAATGCGCCAAAGCTTTCAGAGAGAAAGCTTATCTCTCCGGTGAATGCC 1260
QY 1261 ATCAAAATGGAGATCTCGCGCATCTCAGAAAGGCTGATGTTTGAACGCTGCAATTGCT 1320
DB 1261 ATCAAAATGGAGATCTCGCGCATCTCAGAAAGGCTGATGTTTGAACGCTGCAATTGCT 1320
QY 1321 GAAGACTTCAAGCTGCTCAGAGGGTATTTGTACGCGTTGGGCAATTGCGACGCGGCG 1380
DB 1321 GAAGACTTCAAGCTGCTCAGAGGGTATTTGTACGCGTTGGGCAATTGCGACGCGGCG 1380
QY 1381 GTTCTGGAAGCGGCTCTTACGCTCTGAGAGTATGTTGCTCTGATCTGGAATGCG 1440
DB 1381 GTTCTGGAAGCGGCTCTTACGCTCTGAGAGTATGTTGCTCTGATCTGGAATGCG 1440
QY 1441 CTTGGAATGCTCGTGTTCGCGCTTCTGCACTGCGGCTTGTGCGGGCTAGGAAA 1500
DB 1441 CTTGGAATGCTCGTGTTCGCGCTTCTGCACTGCGGCTTGTGCGGGCTAGGAAA 1500
QY 1501 GAGGCGTCGAGCGCGAGGTCCTTGCAGTGAAGCGGCTGCGGCTGCTGACTGCG 1560
DB 1501 GAGGCGTCGAGCGCGAGGTCCTTGCAGTGAAGCGGCTGCGGCTGCTGACTGCG 1560
QY 1561 CTCAAAAGACTCAATCGAAGAGCACTGCGCAATGCGATCGCATATGTTGGGTGCGCTC 1620
DB 1561 CTCAAAAGACTCAATCGAAGAGCACTGCGCAATGCGATCGCATATGTTGGGTGCGCTC 1620
QY 1621 CTGATACGCGCGCGCTGATTAAGGCGAGGCTCACTGAACAAGGCTCGATCAACAG 1680
DB 1621 CTGATACGCGCGCGCTGATTAAGGCGAGGCTCACTGAACAAGGCTCGATCAACAG 1680
QY 1681 CGGCTGTTTGAATGCGGCTGCGCAAAATTGATGCGCTGTATGCTGTGAAGATCAA 1740
DB 1681 CGGCTGTTTGAATGCGGCTGCGCAAAATTGATGCGCTGTATGCTGTGAAGATCAA 1740

Qy 1741 TCCATGCTGCTGACGAGGCCACACTGTGA 1770
Db 1741 TCCATGCTGCTGACGAGGCCACACTGTGA 1770

RESULT 3

US-08-976-063C-1
Sequence 1, Application US/08976063C
Publication No. US20020182697A1

GENERAL INFORMATION:

APPLICANT: Alexander Steinduechel, Horst Priefert, Jurgen Rabenhorst
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
TITLE OF INVENTION: CONFERYL ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN A
TITLE OF INVENTION: ACID AND THEIR USE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: HP VECTRA
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976, 063C
FILING DATE: 21-NOV-1997
CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 196 49 655.1 (Germany)
FILING DATE: 29-NOV-1996
ATTORNEY/AGENT INFORMATION:

NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 32679 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Pseudomonas sp.

STRAIN: HR199

FEATURE:

NAME/KEY: CDS

LOCATION: 3146..3997

OTHER INFORMATION: /gene="ORF1"

US-08-976-063C-1

Query Match 100.0%; Score 1770; DB 8; Length 32679;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGTCTCTGAGAGGCTTCTTCCGCGTGAATTCTTGAGCGTCTCGACAT 60
Db 21288 ATGCGTCTCTGAGAGGCTTCTTCCGCGTGAATTCTTGAGCGTCTCGACAT 21347
Qy 61 TGGGCTTAGACCCGCTCAGAAACAACCTGCGTTCGCGAGGGGCGCAAAATGGGGAATGG 120
Db 21348 TGGGCTTAGACCCGCTCAGAAACAACCTGCGTTCGCGAGGGGCGCAAAATGGGGAATGG 21407

Qy 121 CGTCGTATCAGCTACGCGGAATGTTCCACAAAGTCGCGCATTCGACAGACTTGT 180
Db 21408 CGTCGTATCAGCTACGCGGAATGTTCCACAAAGTCGCGCATTCGACAGACTTGT 21467
Qy 181 CCTTACGGAATATCGGACAGAGCGTCCGCTCTTATGCTCTTGGAATATGCTTGGAACAT 240
Db 21468 CCTTACGGAATATCGGACAGAGCGTCCGCTCTTATGCTCTTGGAATATGCTTGGAACAT 21527
Qy 241 CTTACAGTGGCATTTTGGGCGTATGTAATGCGGGGATTCCTTATTTGCCCGGTCTCTGCT 300
Db 21528 CTTACAGTGGCATTTTGGGCGTATGTAATGCGGGGATTCCTTATTTGCCCGGTCTCTGCT 21587
Qy 301 TATTCATGCTGTGCGAAGATTTGGCGAAGCTGCTACATGCTAGCTTCTTGCAACCG 360
Db 21588 TATTCATGCTGTGCGAAGATTTGGCGAAGCTGCTACATGCTAGCTTCTTGCAACCG 21647
Qy 361 GGACTGCTCTTTCCTGCGCATATGAGACACTTTCACAGCGCAATTTAGACATTTGCGCG 420
Db 21648 GGACTGCTCTTTCCTGCGCATATGAGACACTTTCACAGCGCGCAATTTAGACATTTGCGCG 21707
Qy 421 GACGAGCTGCGCGCAATCTTCACTCGAGGCGAATTTGCGCGGCGCGCAAGTGTGTT 480
Db 21708 GACGAGCTGCGCGCAATCTTCACTCGAGGCGAATTTGCGCGGCGCGCAAGTGTGTT 21767
Qy 481 GACAGCTGTGAGAGAGCTGTGCGAATTTAGAGAGATTAATGCTTTGCGGCACTGAGC 540
Db 21768 GACAGCTGTGAGAGAGCTGTGCGAATTTAGAGAGATTAATGCTTTGCGGCACTGAGC 21827
Qy 541 CCGGATACGATTTGCGCAAGTCTTGTTCACCTTGTGCTCTTACCAAACTGCTTAAAGCGGTG 600
Db 21828 CCGGATACGATTTGCGCAAGTCTTGTTCACCTTGTGCTCTTACCAAACTGCTTAAAGCGGTG 21887
Qy 601 CCGACTACTGAGGGAATGCTCTGCGCAATCAGACAGATGCTTGTGCAAACTTTCGCGGT 660
Db 21888 CCGACTACTGAGGGAATGCTCTGCGCAATCAGACAGATGCTTGTGCAAACTTTCGCGGT 21947
Qy 661 TTTGTGAGAGAGCGCGGTGCTGTGAGACTGTTCGCTGGAACCAACCTTTCGCGGCGC 720
Db 21948 TTTGTGAGAGAGCGCGGTGCTGTGAGACTGTTCGCTGGAACCAACCTTTCGCGGCGC 22007
Qy 721 AGCCACAACATCGGCAATCGTGTGTAACAAGCGGCGACGTAACCTTGAAGCGAGTAA 780
Db 22008 AGCCACAACATCGGCAATCGTGTGTAACAAGCGGCGACGTAACCTTGAAGCGAGTAA 22067
Qy 781 CCAACGCGCAAGGCTTCCGCGAGAGCTTTCGCACTTTCGCAAACTTTCGCACTGCG 840
Db 22068 CCAACGCGCAAGGCTTCCGCGAGAGCTTTCGCACTTTCGCAAACTTTCGCACTGCG 22127
Qy 841 TACCTGACTGTGCGAAGAGCTGGAAGATTAATGAGGTGCTCTTGAAGCGACAGTACC 900
Db 22128 TACCTGACTGTGCGAAGAGCTGGAAGATTAATGAGGTGCTCTTGAAGCGACAGTACC 22187
Qy 901 CTGCGCGAAGCGCTTCTGCGTGCATGAAGCTGTTCTTTCGCGCGGCTGAGGTGTG 960
Db 22188 CTGCGCGAAGCGCTTCTGCGTGCATGAAGCTGTTCTTTCGCGCGGCTGAGGTGTG 22247
Qy 961 CAAGGATCTGAGATGTTTGAACCGGCTGCTGAACAGCACTGTGTGAGCGCATTTGCG 1020
Db 22248 CAAGGATCTGAGATGTTTGAACCGGCTGCTGAACAGCACTGTGTGAGCGCATTTGCG 22307
Qy 1021 ATGATGCGCGGTCTGAGCATGAGAGACTGCTCTTCTGCACTTTTACACCGGACCG 1080
Db 22308 ATGATGCGCGGTCTGAGCATGAGAGACTGCTCTTCTGCACTTTTACACCGGACCG 22367
Qy 1081 CTGTCGATGCTGTGATTAATGAGGCTGCGAGCGCTGCGAGAGGCAAGCTCGTTCCG 1140
Db 22368 CTGTCGATGCTGTGATTAATGAGGCTGCGAGCGCTGCGAGAGGCAAGCTCGTTCCG 22427
Qy 1141 GTGATGAGAAATTTGAAGGCGCTTTCATGCTGCGCACTCATGAGCGGCTACTGCGCT 1200
Db 22428 GTGATGAGAAATTTGAAGGCGCTTTCATGCTGCGCACTCATGAGCGGCTACTGCGCT 22487
Qy 1201 GCTCTGAACAAATGCGCAAGCTTTCGAGAGGAAGCTATTATGCTCTCGGTATGCC 1260

Db 22488 GCTCTGGAACAAATGCTCCAAAGGTTGACAGAGGAAGCTATTACTCTCGGGTAGGCC 22547
Qy 1261 ATCAAAATGGCAGATCTCTCCGATCTCAGAAAGCTGATGTTTGACGGTCGAATTGCT 1320
Db 22548 ATCAAAATGGCAGATCTCTCCGATCTCAGAAAGCTGATGTTTGACGGTCGAATTGCT 22607
Qy 1321 GAGACTTCAAGGTCCTCAGAGGGATATTGTACAGCGCTTGAGGCAATGGCCACGCGGGCG 1380
Db 22608 GAGACTTCAAGGTCCTCAGAGGGATATTGTACAGCGCTTGAGGCAATGGCCACGCGGGCG 22667
Qy 1381 GTTCTGAAAGCGGCTCTTAAGTCTGAGAGTGTGCTGCTCTGATTCGTGAATGC 1440
Db 22668 GTTCTGAAAGCGGCTCTTAAGTCTGAGAGTGTGCTGCTCTGATTCGTGAATGC 22727
Qy 1441 CTGGAATGCTGCTGTTTCCGCGCTTCTGACCTGCGGCTTGTGCGGAGTGA 1500
Db 22728 CTGGAATGCTGCTGTTTCCGCGCTTCTGACCTGCGGCTTGTGCGGAGTGA 22787
Qy 1501 GAGGCGTCCGACGCGGAGGTCCTTCCAGTGAAGCGGCTTGCGGCTTGCTGACTGG 1560
Db 22788 GAGGCGTCCGACGCGGAGGTCCTTCCAGTGAAGCGGCTTGCGGCTTGCTGACTGG 22847
Qy 1561 CTCGAAAGCTCATCGAGAGCACTGCGAATGCGCATCATGTGAGGCTC 1620
Db 22848 CTCGAAAGCTCATCGAGAGCACTGCGAATGCGCATCATGTGAGGCTC 22907
Qy 1621 CTCGATACGCGCGCGCTGATTAATTAAGGCGAGGCTCATGACAGGCTGATCAACAG 1680
Db 22908 CTCGATACGCGCGCGCTGATTAATTAAGGCGAGGCTCATGACAGGCTGATCAACAG 22967
Qy 1681 CGGCGCTGTTTTCGATAGGCGGCTGCGGAGGATGATGCGCTGATGCTGGAAGATCAA 1740
Db 22968 CGGCGCTGTTTTCGATAGGCGGCTGCGGAGGATGATGCGCTGATGCTGGAAGATCAA 23027
Qy 1741 TCCATGCTGCTGACGAGGCGCACACTGTGA 1770
Db 23028 TCCATGCTGCTGACGAGGCGCACACTGTGA 23057

RESULT 4
US-09-750-986d-1
; Sequence 1, Application US/09750986D
; Publication No. US20030228670A1
; GENERAL INFORMATION:
; APPLICANT: Steinduchel, Alexander
; APPLICANT: Pfeleert, Horst
; APPLICANT: Rabenhofst, Jurgen
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONFERYL
; TITLE OF INVENTION: ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; FILE REFERENCE: VANILLIC ACID AND THEIR USE
; CURRENT APPLICATION NUMBER: US/09/750,986D
; PRIOR FILING DATE: 2000-12-28
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 32679
; TYPE: DNA
; ORGANISM: Pseudomonas sp.
; NAME/KEY: CDS
; LOCATION: (3146) .. (3997)
; OTHER INFORMATION: gene = "ORF1"
US-09-750-986d-1

Query Match 100.0%; Score 1770; DB 10; Length 32679;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGTTCTCTGAGGCGCTTCTCCCTCCGGGTCGAATTCTTGAGCGCTTCGAGCAT 60

Db 21288 ATGCGTTCTCTGAGGCGCTTCTCCCTCCGGGTCGAATTCTTGAGCGCTTCGAGCAT 21347
Qy 61 TGGGCTTAAGACCCGCTCCAGAACAAACTGCTGTGCTGCGAGGCGCGCAATGGGGAATGG 120
Db 21348 TGGGCTTAAGACCCGCTCCAGAACAAACTGCTGTGCTGCGAGGCGCGCAATGGGGAATGG 21407
Qy 121 CGTGTATCAGCTACCGGGAATGTTTCACAAAGTCCGCGCATTCGACAGAGCTTGCTT 180
Db 21408 CGTGTATCAGCTACCGGGAATGTTTCACAAAGTCCGCGCATTCGACAGAGCTTGCTT 21467
Qy 181 CTTAAGGATATAGGAGAGCGCTCCGCTGCTTATGCTCTGGAATGACCTGGAACAT 240
Db 21468 CTTAAGGATATAGGAGAGCGCTCCGCTGCTTATGCTCTGGAATGACCTGGAACAT 21527
Qy 241 CTTAAGGATATAGGAGAGCTTATGATGCGGAGATTCCTTATGCTCCGAGTCTCTGCT 300
Db 21528 CTTAAGGATATAGGAGAGCTTATGATGCGGAGATTCCTTATGCTCCGAGTCTCTGCT 21587
Qy 301 TATTACCTGCTGTCGGAAGATTGGGGAAGCTGCTCAGATCGTAGGCTTCTGCAACCG 360
Db 21588 TATTACCTGCTGTCGGAAGATTGGGGAAGCTGCTCAGATCGTAGGCTTCTGCAACCG 21647
Qy 361 GGACTGCTTTCGCTGCGAGTGCAGACCTTTCAGCGCGCAATTGAGACATTGCGCG 420
Db 21648 GGACTGCTTTCGCTGCGAGTGCAGACCTTTCAGCGCGCAATTGAGACATTGCGCG 21707
Qy 421 GACGAGCTGCGCGCAATCTTCACTCGAGGCGAATTGGCGGCGGCGGACAGGATGTTT 480
Db 21708 GACGAGCTGCGCGCAATCTTCACTCGAGGCGAATTGGCGGCGGCGGACAGGATGTTT 21767
Qy 481 GACGAGCTGCTGAGAGAGCTGCTGAGGATTAAGGAGATTAATGCTTGGGCAACTGGC 540
Db 21768 GACGAGCTGCTGAGAGAGCTGCTGAGGATTAAGGAGATTAATGCTTGGGCAACTGGC 21827
Qy 541 CCGGATACGATTCGCAAGTCTTGTTCATTCTGAGCTTCAACAACTGCTAAAGCGGTG 600
Db 21828 CCGGATACGATTCGCAAGTCTTGTTCATTCTGAGCTTCAACAACTGCTAAAGCGGTG 21887
Qy 601 CCGACTATCTAGGGAATGCTCTGCGGCAATCAGCAGATGCTTCTGCAAACTTTCGCGGT 660
Db 21888 CCGACTATCTAGGGAATGCTCTGCGGCAATCAGCAGATGCTTCTGCAAACTTTCGCGGT 21947
Qy 661 TTTGGTGAAGAGCGCGGCTGCTGAGTGTGAGTGTGCGGCGGGAACCACTTCGCGGCG 720
Db 21948 TTTGGTGAAGAGCGCGGCTGCTGAGTGTGAGTGTGCGGCGGGAACCACTTCGCGGCG 22007
Qy 721 AGCCACAACATCCGCAATCGTGTGTACAAAGCGGCACTTACTTACGACGCTTAA 780
Db 22008 AGCCACAACATCCGCAATCGTGTGTACAAAGCGGCACTTACTTACGACGCTTAA 22067
Qy 781 CCAACCGCCCAAGGCTTCGCGAGAGCTTCGCACTTGAGGCAATCTTCCACTGCG 840
Db 22068 CCAACCGCCCAAGGCTTCGCGAGAGCTTCGCACTTGAGGCAATCTTCCACTGCG 22127
Qy 841 TACCTACGTGCGGAAGCTGAGGAAATTAGTGGTGCCTTGAAGGAGACAGTACC 900
Db 22128 TACCTACGTGCGGAAGCTGAGGAAATTAGTGGTGCCTTGAAGGAGACAGTACC 22187
Qy 901 CTGCGGCAACGCTTCTTCCGCTGCAATGAAGCTTCTTCTTCCGCGGCTGAGGTTGTCG 960
Db 22188 CTGCGGCAACGCTTCTTCCGCTGCAATGAAGCTTCTTCTTCCGCGGCTGAGGTTGTCG 22247
Qy 961 CAAGGATCTGGGATGCTTGGACCGGCTGCTGAACAGCATGCTGTGAGCGCATTCGCG 1020
Db 22248 CAAGGATCTGGGATGCTTGGGATGCTTGGGATGCTTGGGATGCTTGGGATGCTTGGG 22307
Qy 1021 ATGATGCGGAGTCTGGGATGAGAGAGCTGCTTCTGCACTTTCACACGAGACCG 1080
Db 22308 ATGATGCGGAGTCTGGGATGAGAGAGCTGCTTCTGCACTTTCACACGAGACCG 22367
Qy 1081 CTGTCATGAGCTGCTGATTAATGGGCTGCAAGCGCTGCTGAGAGTCAAGCTGTTCCG 1140

Db 22368 CTGTCGATGCTGCTTACATTGGGCTGCCAGCCCTGCTGCGAGGTCACAGTCGTTCCG 22427
Qy 1141 GTTCGATGGGAAATTTGAAGGGGGTTCCATGTCGCCACGTCAATGAGCGGCTACGCGCT 1200
Db 22428 GTGCGAGGAAATTTGAAGGGGGTTCCATGTCGCCACGTCAATGAGCGGCTACGCGCT 22487
Qy 1201 GCTCTGAAACAAATATCCCAAGGCTTCGACGAGGAGGCTATTACTCTCCGCTGATGCG 1260
Db 22488 GCTCTGAAACAAATATCCCAAGGCTTCGACGAGGAGGCTATTACTCTCCGCTGATGCG 22547
Qy 1261 ATCAAAATTTGGCAGATCTCTCCCATCTCTCAAGAAAGCTGATGTTTGAACGGTCAATTTGCT 1320
Db 22548 ATCAAAATTTGGCAGATCTCTCCCATCTCTCAAGAAAGCTGATGTTTGAACGGTCAATTTGCT 22607
Qy 1321 GAAGACTTCAAGCTCTCTCAAGGAGTATTGTACAGGTTGGGCGCATTTGCGACGCGGCG 1380
Db 22608 GAAGACTTCAAGCTCTCTCAAGGAGTATTGTACAGGTTGGGCGCATTTGCGACGCGGCG 22667
Qy 1381 GTTCTGGAAGCGGCTCTTACGTCCTGACGTAAGTGTGCTGCTCTGATCTGTAATGC 1440
Db 22668 GTTCTGGAAGCGGCTCTTACGTCCTGACGTAAGTGTGCTGCTCTGATCTGTAATGC 22727
Qy 1441 CTTCGATGCTGCTGTTTCCGCTCTTCTGCACTGCCGCTTGTGCGGCTAGAGAAA 1500
Db 22728 CTTCGATGCTGCTGTTTCCGCTCTTCTGCACTGCCGCTTGTGCGGCTAGAGAAA 22787
Qy 1501 GAGGCGTCGAGCGCCGAGGCTTTCAGTGCAGTGCAGGCTTCCGAGCTGTTTGTGTAATGC 1560
Db 22788 GAGGCGTCGAGCGCCGAGGCTTTCAGTGCAGTGCAGGCTTCCGAGCTGTTTGTGTAATGC 22847
Qy 1561 CTCGAAACGACTCAATGAGAAAGCACTGGCAATGCCATGTCATCATGTGGTAGGAGCTC 1620
Db 22848 CTCGAAACGACTCAATGAGAAAGCACTGGCAATGCCATGTCATCATGTGGTAGGAGCTC 22907
Qy 1621 CTCGATACGCGCCGCTGATGTTGAAGGGGAGAGTCACTGACAAAGGCTTCGATCAACCG 1680
Db 22908 CTCGATACGCGCCGCTGATGTTGAAGGGGAGAGTCACTGACAAAGGCTTCGATCAACCG 22967
Qy 1681 CGGCGCTGTTTTCGATGAGCGGCTGGGCGAAAGTGTATGCTGTGTGTAAGATCAA 1740
Db 22968 CGGCGCTGTTTTCGATGAGCGGCTGGGCGAAAGTGTATGCTGTGTGTAAGATCAA 23027
Qy 1741 TCCATGCTGCTGTCGAGGAGGCGACACTGTGA 1770
Db 23028 TCCATGCTGCTGTCGAGGAGGCGACACTGTGA 23057

RESULT 5
US-10-369-493-44177
; Sequence 44177, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44177
; LENGTH: 1725
; TYPE: DNA
; ORGANISM: Rhodospseudomonas palustris
US-10-369-493-44177

Query Match 29.8%; Score 527.8; DB 15; Length 1725;
Best Local Similarity 56.6%; Pred. No. 4e-155;

Matches 976; Conservative 0; Mismatches 747; Indels 0; Gaps 0;
Qy 2 TGGGTTCTTCGAGGCGCTTCTCCCTCCCGGAGTGAATTTCTGAGCGTCTGACATTT 61
Db 2 TGGGCTGACCAACACGCTTACCGACTATCCGGTGGGATTACGACCAGGCTGACACT 61
Qy 62 GGGCTAAGACCCGTCGAAACAACTGGCTTCTGCGAGGCGGCGAAATGGGAAATGCG 121
Db 62 TGGCCGAGACGGGCGCCGACCGGGTGTTCATGGCCGAGCGGAACGGATGAGGGGCGTGGC 121
Qy 122 GTTCGATTCAGCTACGGGAAATGTTCCAAAGTCCGCGCCATTCGACAGAGCTTCTTC 181
Db 122 GGAGATTCAGCTATGCGGAGATGTCGGCGCGGACACATGCTCGCGCTATATG 181
Qy 182 CTTACGAGCTATTCGCGAGAGCTGCGCTCTTATGCTCTGAGAAATGACTGGAACATC 241
Db 182 CGGCGGACTGTGCGCCGAAACGGCGGATATGATTTCTGCGGTAATTCGATGACATG 241
Qy 242 TTCAGCTGCAATTTGGGGTATGATGCGGGCAATTCCTATTGCGCGGTCTCTGCTT 301
Db 242 CGATGGTGAATGTTCCGCGCGCTGTATGCGGGCGTGGCAATGTCGGTGTGCGCGCTG 301
Qy 302 ATTCACTGCTGTGCGAAGATTTGGCGAAGCTGCTGTCATCTGTAAGTCTTTGCAACCG 361
Db 302 ATTGCTGTGTGTCGAGATTAAGGCAATGCGCAATGCTGCTGCTGCTGCTGCTGCTG 361
Qy 362 GACTGGTCTTTCGCGGATGCGAGCACTTTCAGCGCGCAATTCGACCATTTGCGCG 421
Db 362 GGTGTATCTTTCGCGGATGCGAGCACTTTCGCGCGCAATTCGACCATTTGCGCG 421
Qy 422 ACGACGTGCGCGCAATCTTCTACTGAGGCGAATTTGGCGGGCGGCGACGTTGATTTG 481
Db 422 AGGATGTGCAAGCTGCTGCGCAACGCGGCGAGGAGTGAAGGGGCGCAAGTGAATGCTG 481
Qy 482 ACAAGCTGTGAGAGAGCGCTGTGGGATGAGGCAATATGCTTTGGGCGCAACTGGCG 541
Db 482 CGGAATGTGTGGAGCGCGGAAACATCCGAATCTGCGCGCAAGCAAGAGGATGAGCG 541
Qy 542 CGGATGCAATTCGCAAGTCTTGTTCATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 601
Db 542 ACGACACCAATGCTGTAAGTTTTGTGTCGTCGCGGATTCGACCGGCAATTCGAGGCG 601
Qy 602 CGAATCTGAGGCAATGCTTGTGCGGCAATGAGCAATGCTTTCGCAACTTTCGCGTTT 661
Db 602 TCAATGCGAGCGAGATGATCTGCGCAATCAAGTATGATTCGCGAGGAGTGCCTTCC 661
Qy 662 TTGTGTAAGAGCGCGCGGTGCTGTGAGACTGTGCGCGGAAACACACTTCGCGGGA 721
Db 662 TGAAGACGAGCGCGCGGTGATGCTGAGCTGTGCGCGGAAATCACACTTCGCGGGA 721
Qy 722 GCCACAACTGCGCATGCTGTGTACAAAGCGCGGCACTTACTTGAACGAGTAAAC 781
Db 722 ACCACAAATTCGCTGAGCGCTGTTCAAGCGGCGCTCATGTAATTCAGACGAGCAAC 781
Qy 782 CAACGCGCCCAAGGCTTGCAGAGAGCTTTCGCACTTGAAGCGAAATCTTCCACTGCGT 841
Db 782 CCAACGCGCGCGGAGATGCTTCCACCAATTCGCGCAATTCGCGCGGAGAGTGCCTGT 841
Qy 842 ACTCACTTGTGCGAAGAGCTGAGAGATGATGTGTCCTTGAAGGAGACAGTACC 901
Db 842 ATTTCAACGTTCCGAAGGCTACAGAGTCTCTGCGGTGCTGCGGAAAGACAGACAGT 901
Qy 902 TGGCGCAAGCTTCTTCGCTGCATGAAGCTTCTTCTTCGCGCGGCTGAGTTCGCG 961
Db 902 TGGCGCAAGCTTCTTCGCTGCATGAAGCTTCTTCTTCGCGCGGCTGAGTTCGCG 961
Qy 962 AAGGATCTGGGATGCTTTGAGACCGGCTGCTGAAACAGACTGTGTGAGCGCATTCGCA 1021
Db 962 CGCATGTCTGAAACGGGCTTGAAGAGTGCCTGCGGAGACCGGCGCGGCTGCGCA 1021
Qy 1022 TGATGCGGCTGCTGCGCATGAGAGAGATGCTCTTCTTCGCACTTTTACACGAGCGCG 1081
Db 1022 TGCTACCGGCTTGGCGGCGACCGAGACCGCGCGTTCCTTCATGTCGTGAGACGCGCA 1081

QY 1082 TGTGATGGCTGTGTTACATTTGGGCTGCGAGCGCTGGCTGCGAGTCAAGCTGTTCCGG 1141
DB 1082 CCAGTGTCTCGGCAATGTGGGCTGCGGATGCCGGCAACGAGGCAACAGTGGTCCGA 1141
QY 1142 TCGATGGGAAATTTGGAAGGGGCTTTCCATGATCCGCACTCATAGCGGCTACTGCGGTG 1201
DB 1142 ACAACGGCAAGCTCGAAGTCCGGCCCAAGGGGCGCAACATCACCCCGGCTATTGGGCGG 1201
QY 1202 CTCTTGAACAAATGCCCCAAGCGTTTCAGAGAGGCTATTACTGCTCCGGTGAATCCA 1261
DB 1202 CGCCGAGCTGACCGATTAAGCGCTTCGACGAGAGGGGCTTCTAACAGCTCAAGATGCGC 1261
QY 1262 TCATATTTGAGATCTCTCGCGGATCTTCAGAAAGTCTGATGTTTGAACGTTGAATTCG 1321
DB 1262 TGAAGCGGCTGATGCGCAAGACCTTTCGCGGCTTCGATTTGAGAGGCGGAGTCTCGG 1321
QY 1322 AAGACTTCAAGCTGTCTCTCAAGGGGTAATTTGTCAAGCGTTGGGCAATTCGCAACGCGGCG 1381
DB 1322 AAGACTTCAAGCTGTCTCTCAAGGGGTAATTTGTCAAGCGTTGGGCAATTCGCAACGCGGCG 1381
QY 1382 TTCTGGAAGGCGGCTCTTACGTCCTGAGAGTAAGTGTGCTGCTCTGATCGTGAATGCC 1441
DB 1382 TTGCGGCTGCGCTCTCTGCTGCGGAGCGTGGTGAATCGCGGCTTCGACCGGATTAACG 1441
QY 1442 TTGGAATGCTCGGTTTCCGCGCTTCTGCACTGCGCGTGGCTGTGCGGGGCTAGAAAAG 1501
DB 1442 TCACCGCGCTGGGATCTCTGATCCGACGCGCTGCAAGTGAATGCAACGCTGCCG 1501
QY 1502 AGCGCTGCGACGCGGAGTCTTGCAGTGAAGCGGCTGGGCTGGTGTGCTGATCGTGC 1561
DB 1502 TGGAAACCTCGCGCGGATGGCGGCGCAACATCTGATCCCGAGCGCTTCGCGAGGCT 1561
QY 1562 TCAGAGACTCAATCGAGAGCACTGGAATGCCACTGCATCATGTGGGTAGGCTCC 1621
DB 1562 TCGCAGCGCTGCTGACGAGCGGCTGCTCAACCGGCTCACCGCGCGCTGGCTGC 1621
QY 1622 TCGATACGCGCGCTGATGATTAAGGCGGAGTCACTGCAAGGGCTGATCAACAGC 1681
DB 1622 TCGGCAACCGCTGATGATGAGAGGAGATCAACGAGGCTCGGTCAACAGC 1681
QY 1682 GCGCTGTTTGCATGCGGCTCGGCGAAGTGTATGCGCTGTA 1724
DB 1682 GGGCGCTGCTGGAATATCGGCGCTCGTGTATCGCGATCTTTA 1724

RESULT 6

US-10-369-493-35384
; Sequence 35384, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 35384
; TYPE: DNA
; LENGTH: 1728
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-35384

Query Match 22.5%; Score 398.4; DB 15; Length 1728;
Best Local Similarity 52.8%; Pred. No. 2,2e-114;
Matches 908; Conservative 0; Mismatches 806; Indels 6; Gaps 2;

QY 13 GAGGCGCTTCTTCCCTTCCCGGCTCGAATTTCTGAGCGTCTGAGGATGGGCTAGAAC 72
DB 1 GATCCGCTTGGCCCCCTATCCGAGAAATGATAGAGTCTCTTCACTGATGCTTCC 60
QY 73 CGTCCAGAACAAACCTGCGTTGTGCGCAGGCGGCAATGGGAAATGGCTGTATCAGC 132
DB 61 GCACCGGAGCGCAATGATGGCGAACGTCAGAGTCCGAGGCAATGGGCGGAGTACG 120
QY 133 TACGCGAAATGTTCCAAACGTCGCGGCAATGCAACAGCTTTCCTTACGACTA 192
DB 121 TACGCGAGGCTCTCCAGAAATACGCGATAGGGCAGTTCCTTCGATCATGACTT 180
QY 193 TCGGCAAGGCTGCGGCTTATCGTCTGAAATGACCTGGAAATCTTCACTGAGCA 252
DB 181 TCGGTGAGAGGCGCGCTGCTGCTTCTGAAATTTCAATCGAGCATGCTGATGATG 240
QY 253 TTGGGGCTATATGATCGGGCATTCCTTATGCGCGGATCTCTGTTATTCATGCTG 312
DB 241 CTGGGCGCGAGATGCTGGTATCGGCTCGGCTGCAATCAGCGCGGCTATGCG---CAGG 297
QY 313 TCGCAAGATTTGGCGAAGCTGCTCAATCGTATGTTCTTGCACCGGAGCTGTCTT 372
DB 298 TCGCGCATCTCAACAACTCGCGCAATCCGCGGAGATCAACCCGCGGATGTTT 357
QY 373 GCTGCGGAGCAGACCTTTCCAGCGCGCAATTGAGACCACTTGGCGGAGCGTGGCC 432
DB 358 GCGGAGATGCGAGCGCTTTCGCGGAGCTGCGGAGGTTTTCAGACGAGTACACG 417
QY 433 GCAATCTTCACTGAGGCGAATTTGGCGGCGGCGC---ACGCTGATTTTGAAGCTG 489
DB 418 CTGATGGGCTCGCGCAACCTCGCGGAAGATCGTTCAATATCGTTTCAATTTGAAACCTG 477
QY 490 CTGAGAGCTGCTGAGGATTTGAGCAATATGCTTTGCGCAATGCGGCGGCTGATG 549
DB 478 CTGGAACGAGCGGACCGAAGCGTGTGATCGCTTTCGATCGGCTGCGGCGGATACC 537
QY 550 ATTGCGAAGTCTGTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 609
DB 538 GTGCGCAAAATCTTCTTTCACGTAGAGAGACGACTGTTTGGCGCAAGGGCTATCAGACG 597
QY 610 CAGCGATGCTGCTGCGCCATCAGAGATGCTTCTGCAAACTTTCGCGGTTTGTGTA 669
DB 598 CAGCGATGTTGCTGCTCAATCAGAAATGATTTGCCGATTTGATGCTACTTTCGAG 657
QY 670 GAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 729
DB 658 GAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 717
QY 730 ATGCGATGCTGTTGTAACAGCGCGGCACTACTTACCTTGAAGAGCGTAACCAACGCG 789
DB 718 TTCAACCTGCTCTATATATGCGGTACTTATATATGATGCGGCAAGCCAGCGCG 777
QY 790 CAGGATTCGCGCAGAGCGCTTCCCACTTGAAGGAAATCTTCCACTGCGTACT 849
DB 778 GCGCAATGCGCAGAGCACTGAGCAATCTCCGGGCAATATCGCCACCTGATTTCAAC 837
QY 850 GTGCGAAGGCTGAGAGATTAATGAGTGCCTTGAAGCAGACATACCTGCGGAA 909
DB 838 GTTCCCGCTGCGCAGAAATGCTGCTGCAAGGCAATCGCAAGAGCAGGCGCTGCGCG 897
QY 910 CGCTTCTGCGCTGCGAATGAGTGTCTTCTGCGCGCGCTGAGTGTGCGAAGGATC 969
DB 898 AGCTTTTTCGCGACTGAAAGATGCTGATGATGCGGCTGCGCAATGCGGAGCAGACACA 957
QY 970 TGGGATGTTTGAACCGGCTGCTGGAACAGCACTGTGTGAGCGCATTTGCGATGATGCG 1029
DB 958 TGGGATGCTGCGAGCGAGCTTTTCATGCGCAACGCTGGCGATTCGATGGGATGCC 1017
QY 1030 GGTCTGGGATGACGAGAGCTGCTCTTCTGCACTTTTACACCGGAGCGCTGTGATG 1089
DB 1018 GGCTCGGCTCGACGAGAGCGGACCCCTTCTGCTGCTTTCGACCGAACCAGCGAGACAG 1077

QY 1090 GCTGTTAATTGGCTGCGGCGCTGCGAGTCAAGCTGTTCCGATCGATGGG 1149
DB 1078 CCGGGCAATATCGGTATCCGGGCGAGGGCGTGAACATGAACTCGGCCCTTGATGGC 1137
QY 1150 AAATTGGAAGGGGCTTTCATGGTCCGACAGTCAATGAGCGGCTTACCTGCTGCTGAA 1209
DB 1138 CGGTACGAATTAAGGCTGAAAGGCGGACATCAAGCCCGGTTATCGGCTTAATGGGAA 1197
QY 1210 CAATATGCCCAAGCTTTCAGAGGAGGCTATTAAGCTCCGGTATGCGATCAATTTG 1269
DB 1198 TTACGCGCGGAGCTTTCAGAGGAGGCTTTCATGCAATCGGCAATCGGCTTAATTT 1257
QY 1270 GCGATCTGCGGATCTCAGAAAGGCTGATGTTGACGCTGCAATTCGTGAAGCTTC 1329
DB 1258 GCGGTGCGGAGCATCGGCGCGGCTTATTTGATGGCGCATGGGGAATCTTC 1317
QY 1330 AAGCTGCTCAGGGGATTTGTCAGGCTTGGGCTATGCGACGCGGCGGCTTCTGAA 1389
DB 1318 AAGCTGCAAGCCGACCTGAGGCTGCGGCTTGGGCTTGGCGGCAATTAATCAATG 1377
QY 1390 GCGGCGCTTACGTCCTGAGAGTATGTTGCTGCTCTGATGATGCTTGGATTTG 1449
DB 1378 TTGCGCGGCGGTATCCGTATGCTGCTATACGAGGCAAAACCGGCGCACTGGGCGCT 1437
QY 1450 CTGCTGTTTCCGCGCTTCTGCACTGCGGCTGCTTGTGCGGCGTAAAGAAAGCGCTCG 1509
DB 1438 CTGCTGCTGCTATTAACCGGCGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1497
QY 1510 GAGCGGAGGCTGCTTGGCACTGAGCGGCTTGGGCTTGGCTGCTGCTGCTGCTGCTG 1569
DB 1498 GAGCGGAGGCTGCTTGGCACTGAGCGGCTTGGGCTTGGCTGCTGCTGCTGCTGCTG 1557
QY 1570 CTCAATCGAAGCACTGCGCAATGCGATGCGATGAGGCTGCTGCTGCTGCTGCTGCTG 1629
DB 1558 CATCAAAAGAGGCGGAGGCTTCCGCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1617
QY 1630 CCGCGCTGATTAATGAAGGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1689
DB 1618 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1677
QY 1690 TTGCAATGCGGCTGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1729
DB 1678 CTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1717

RESULT 7
US-10-369-493-38427
Sequence 38427, Application US/10369493
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 38427
LENGTH: 1728
TYPE: DNA
ORGANISM: Agrobacterium tumefaciens
US-10-369-493-38427

Query Match 22.5%; Score 398.4; DB 15; Length 1728;
Best Local Similarity 52.8%; Pred. No. 2.2e-114;
Matches 908; Conservative 0; Mismatches 806; Indels 6; Gaps 2;

QY 13 GAGCGCTTCTTCCCTTCCCGGCTGCAATTTCTGAGCGCTGCGAGCTTGGGCTAAGACC 72
DB 1 GATCGCTTGGGCCCTTATCCGAGAAATTAATGAGCGCTTCTTCACTGCTGCTGCTGCTG 60
QY 73 CTTCCAGAACAACTTCCGCTTGTGCTGCGAGGCGGCAATGGGAAATGGGCTGCTATACG 132
DB 61 GCAACCGAGCGCAATGATGAGGCGGAGCTGCAAGGCTGCGAGGCAATGGGCGCGGCTAGC 120
QY 133 TACGCGAAATGTTCCACAACGTCGCGGCTGCAAGAGCTTGGCTTCTTCAAGCACTA 192
DB 121 TACGCGAGGCTTCCACAAGATGCGGCTATGCGGCACTTCTTCCATCATGACCTT 180
QY 193 TCGGAGAGCGTCCGCTTATGCTGCTGGAATGACCTGGAACCTTCACTGCGCA 252
DB 181 TCGTGAGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
QY 253 TTTGGGCTATGATGCGGCGATTCCTTATGCGCGGCTGCTGCTGCTGCTGCTGCTGCTG 312
DB 241 CTTCGCGGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 297
QY 313 TCGCAAGATTTGGCGAAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 372
DB 298 TCGGCGGATCTCAACCAATCTGCGGCAATGCGGCGGCAATACCCCGGAGATGTTT 357
QY 373 GCTGCGGATGCAACCTTTCAGCGGCAATGAGCACTTTCGCGGAGCGAGTGC 432
DB 358 GCGGAGAGTGCAGCGCTTTCGCGGCGCTGCGGAGGTTTTCAGCAGGTTACCG 417
QY 433 GCATCTTCACTGAGGCGAATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 489
DB 418 CTGCTGCGGCTGCGCACTGCGGAGGATGCTTCAATACCTTCAATTCGAAACCTG 477
QY 490 CTGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 549
DB 478 CTGGAACGAGCGGCAACCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 537
QY 550 ATTGCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 609
DB 538 GTGCGCAAAATCTTCTTCAAGTCAAGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 597
QY 610 CAGCGAATGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 669
DB 598 CAGCGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 657
QY 670 GAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 729
DB 658 GAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 717
QY 730 ATGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 789
DB 718 TTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 777
QY 790 CAGGCTTGGCGGAGCGCTTGGCACTTGAAGGAAATCTTCCCACTGCGTCACT 849
DB 778 GCGCAATGCGGCGGAGCTGCAATCTGCGGCAATGCGGCAATGCGGCAATGCGGCAAT 837
QY 850 GTGCGAAGGCTGCGGAGGAAATGAGGCTGCTGCGGCAAGCACTGCGGCAAG 909
DB 838 GTTCCGCGGCGGCAAGAAATGCTGCTGCGGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 897
QY 910 CCGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 969
DB 898 AGCTTTTTCGCGAAGTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 957
QY 970 TGGGATGCTTGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1029
DB 958 TGGGATGCTTGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1017
QY 1030 GGTGCGGATGCGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1089
DB 1018 GGTGCGGATGCGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1077
QY 1090 GCTGTTAATTGGCTGCGGCGCTGCGAGGCTGCGAGGCTGCTGCTGCTGCTGCTGCTG 1149

Db 1078 CCGCGCAATATCGGTATCCCGCGCAGGGCGTACATGAAACTCGTCCCTTCGATGGC 1137

Qy 1150 AAATTGAAAGGCGTTTCCATGTCCTCCGACGTATGAGCGGCTA CTGAGCGTCTTGAA 1209

Db 1138 CGGTACGAATTCACGCTGAAAGGCCGGAACATCACGCCCGGTACTGGCTTAATGAGCGAA 1197

Qy 1210 CAAAATGCCAAGCGTTGCACGAGGAAAGGCTATTACGTCTCCGCTGATNGCATCAATTG 1269

Db 1198 TTGACGCGCGGACGCTTCGACGAGGAAAGGTTCTATGAAATCGCGATACCGTCAAAATT 1257

Qy 1270 GCAGATCTCGCCGATCTCAGAAAGGTTCTGATTTTGAACGCTGAAATTGCTGAAGCTTC 1339

Db 1258 GCGCTGCGGAGATCCGCGCGCGGCTTTATTTGATGAGCGCATGCGGAAAACTTC 1317

Qy 1330 AAGCTGTCTCAGGGGATATTGTCAACGTTGGGCAATTGGGCAACGCGGCGGTTCTGAA 1389

Db 1318 AAGCTGCAGACCGGACCTGGGTGGCGCTTTGTGGCGCAATTAGTCAACATG 1377

Qy 1390 GCGGCTCTTACGCTCTGACAGTAGTGTTGTCTCTCATGCTGATGCTTGAATTG 1449

Db 1378 TTGGCGGCGGTATCCGTATGTCCTGTATACCGGCGAAACCGGCGCAACTGGGCGCT 1437

Qy 1450 CTCGTGTTTCCGCGTCTTCTCGACTGCGCTGTGTGGGGCTAGGAAAGAGCGCTG 1509

Db 1438 CTCGTGTCGCAATTCATACCGGCGCTACGGAACCTGTTGCGGTTGCGCAACATCTCTCC 1497

Qy 1510 GACGCGAGGCTGTTCCAGTGAAGCGGTTCCGGGCTGTTTGCTGACTGCTCAACGA 1569

Db 1498 GATGCGGAGATCATCCGCAACCGAGTGTACAGGCGCAAGTTGTCCGAACTTTCCGCG 1557

Qy 1570 CTCAATCGAAGCACTGCGCAATGCGCATGTGTGGTAGGAGCTCTTCGATACG 1629

Db 1558 CATCAAAAACAGGCGAGCGGTTCCGCAAGCGCGTCACTGCGCAATTCTGTGATGGAAGAC 1617

Qy 1630 CCGCGGTGATGATTAGGGCGGAGTCACTGACAAAGGCTCGATCAACGCGGCGTGT 1689

Db 1618 GCGCTGCGCTTGAGAAAGCGAGGTTACGACAAAGGATCGCAACACGAGCGCGGTG 1677

Qy 1690 TTGCAATGCGGCTCGCGCAAAAGTTGATGCGCTGTATCGT 1729

Db 1678 CTTCTGCATCGCAAGGAGCTCGTAAAGTCACTTTAGCGCG 1717

RESULT 9

US-10-369-493-38007

; Sequence 38007, Application US/10369493

; Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 38007

LENGTH: 1660

TYPE: DNA

ORGANISM: Agrobacterium tumefaciens

US-10-369-493-38007

Db	1	GATCCGCTGGCCCCCTATCCGAGAAATTGAATGAGGCTCTCTTCACTGATGTGCTTCG	60
Oy	73	CGTCCAGAAACAAACCTTCGCTTGTCTGCCAGGCGGCAAAATGGGATAGGCGTTCGATTAGC	132
Db	61	GCACCCGAGGCGACCATGATGGCCGACCGCTCAGGGGTGCCAGGCAATGGCCCGGCTAGC	120
Oy	133	TAGCGGAAATGTTTCCAAAGTCCGGGCGCATCGCACAGAGCTTGCTTCCCTTACGGACTA	192
Db	121	TACGCCAGAGGCTCTCGACCAAGATACGGCGGATACGGCCAGTTCTCTTCGATCATGACTT	180
Oy	193	TCGGCAGAGCGTCCGCTGCTTATGCTCTGGAATGACCTGAAACATCTTCAAGCTGGCA	252
Db	181	TCGGTCCAGAGGCGCGCTGCTGCTCTTCTGAAATTCATTCGAGCATGCGTGGATGCTG	240
Oy	253	TTTGGGCTATGATATGGGGGATATCCCTATTTGCCCGGATCTCTCTGTTATTCACGCTG	312
Db	241	CTTCGGCGCCACGACATGTCGGATGCTCTCGGCTCGGCTGCATCAGCCGGGCTTAGC--	297
Oy	313	TCGCAAGATTGGCGAAGCTGCGTCACATGATGAGTCTTTCGCAACCGGAGACTGCTCTT	372
Db	298	TGGCGCGATCTCACCAATCTGCCGACATCGCGGGGAGATCAACCCGGGGATGGTTTTT	357
Oy	373	GCTGCCGATGCGAGACCTTTTCAGCGCGCAATTGAGACCATTTGCGCGAGACGAGTCC	432
Db	358	GCCAGGATGGAGCGCCCTTTCGCGGGCGCTCGCGAGGTTTTCGACGACGGTACCG	417
Oy	433	GCAATCTTCACTCGAGGCGAATTGGCGGGCGGGC--ACGGTAGATTTTTGACAGCTG	489
Db	418	CTGGTCCGCTCCGGAACCTGCCGGAAGATGTTCCATACGTTTCACTTTCGAAACCTTG	477
Oy	490	CTGAGACGACCTGCTGGGATTTGAGCGAGATATGCTTTTCGGGCACTGGCCCCGATAG	549
Db	478	CTTGAACCGAGCGGACCGAAGCGGTGTGATCGTCTTTCATGCGGTCCGGGCGGATACC	537
Oy	550	ATTGCCAAGTCTTGTTCACTTCGCGCTACCAAACTGGCTAAGGCGGTGCCGACTACT	609
Db	538	GTCGCGAAATCTTTTTCAGTACGAGGACGACTGCTTCCGCCAAGCGGTATCAGACG	597
Oy	610	CAGCGAATGCTCTGCGCCATTCAGCAGATGCTTTCGMAACTTTCGCCGTTTGTGTGAA	669
Db	598	CAAGCAATGTTGCTCCATCAGGAATGATTCGCAATTGTTATGACTTTCGCGAG	657
Oy	670	GAGCGCGCGGTCTGCTGTGACTGTTGCCGTGGAACCAACCTTTCGGCGGACGACCAAC	729
Db	658	GAGCGCGCGGTCTGCTGTGATTGGGCCCATGGAACCAACGCGCGGGGAACAAAGTG	717
Oy	730	ATCGGCATCTGTTGTATCAACGCGGCGACGCTACCTTACGCAAGCTTAACCAACCGCC	789
Db	718	TTCAACTCTGCTCTTATATGGGGTACTATTTATTCATGATCGGCGAAGCCAGGCCG	777
Oy	790	CAAGGTTTCGCGAGACGCTTCGCACTTGAAGCAAACTCTCTCCACTGCGTACTACT	849
Db	778	GCGCAGATCCGCGCAGACATGGAACAACTCCGGGACATATCGGCCACCTGATATTTCAAC	837
Oy	850	GTCGCGAAGGCTGGGAGGAATTATGATGGTGGCCCTTAGGAGAGACAGTACCTCGCGGAA	909
Db	838	GTTCCCGGTGCGACGAAATGCTGTCTCAGGCAATGCGCAAGACGAGGCGCTGTGCCG	897
Oy	910	CGCTTCTTCTGCTGCGATGAAGCTGTCTTCTTTCGCGGCGGCTGGGTGTGCAAGGATC	969
Db	898	AGCTTTTTCGCGACTGAAGATGCTGATGATGCCGGTCCCGGACATGGCGAGACACA	957
Oy	970	TGGGATGTTTGAACCGGCTGCTGGAACAGACTGTGTGTAGCGCACTTCCATGATGCGC	1029
Db	958	TGGGATCGCTGACGAGCTTTCCATGSCAACGCTCGGCATGCGGTTCTCATGAGGTGC	1017

Query Match	21.0%	Score 372;	DB 15;	Length 1660;
Best Local Similarity	52.7%	Pred. No. 4.5e-106;		
Matches 877; Conservative	0;	Mismatches 780;	Indels 7;	Gaps 3;

13 GAGGCGTTCCTCCCTTCCGGGTCGAATCTTGAGCGTCTGAGCATTTGGCTAAGACC 72

Db 1078 CCCGCAATATCGGTATCCCGCGCAGGGCGTGACGATGAACTCGTCCCTTGATGGC 1137

QY 1150 AATGTGAGGGGTTTCATGTCGACGTGATGAGCGCTACGCGTCTCTGAA 1209
DB 1138 CGGTGACAAATTACGGCTGAAAGCCGACATCAGCCGGTTACTGCGAAATGCGGA 1197
QY 1210 CAAATGCCAAGGTTGACGAGAGAGGCTATTACTGCTCCGCTGATGCGCAATATG 1269
DB 1198 TTGACGGCGGACAGCTTTCGACGAGAGAGGTTCTATGAAATCGCGCATCCGTAATTTG 1257
QY 1270 GCGATCTCTCCGATCTCTGAGAAAGTCTGATGTTTGAAGGTGCAATTGCTGAAACTTC 1329
DB 1258 GCGGTGCGGACGATCCGCCCGGCTTTTATTCATGAGCCGCAATGCGGAAACCTTC 1317
QY 1330 AAGCTGCTCAAGGAGATTTGTGAGGTTGGGCAATTGGGCAAGCGGGGCTTCGAA 1389
DB 1318 AAGCTGACACCGGACCTGAGGTGCGGTGCGGCTTTCGCGGCAATTAATGACAATG 1377
QY 1390 GCGGCTCTTACGCTCTGACGTAAGTGTGCTGCTCTGATCGTAATGCTTGATTTG 1449
DB 1378 TTGCGCGGAGTATCGGTGATGCGGTGATTAACGGGCGGAAACCGGGCGCAATGGGCGT 1437
QY 1450 CTGTGTTTCCGGTCTTCTGACTGCGGCTTGTGCGGGCTAGAGAAAGAGCGTTCG 1509
DB 1438 CTGCTGCTGCTATTCATACCGGCGCTACGTGAGCTTGTTCGCGGTTGCAACATCTCTCC 1497
QY 1510 GAGCGGAGGTGCTGCTGCAATGAGCGGTTCCGGCTGTTTGTGCTGACTGCTCAAGCA 1569
DB 1498 GATGCGGAGATCATCCGCGACCCGAGTGTCAAGGCGCAATTTGCGGAACTTTCGCGC 1557
QY 1570 CTCAATCGAAGCAACTGCAATGCGCATGCGATCATGTCGAGGCTCTCGATACG 1629
DB 1558 CATCAAAAACGAGGAGCGGTTCCGAAAGCGGCTATGCGCATTCCTGTGATGGAAGAC 1617
QY 1630 CCGCCTCATTTGTAAGGCGGAGGTCACTGACAGGCGTCGAT 1673
DB 1618 GCGCTG-CGCTTCGAGAGGCGAGGTTACGACAAAGGATCGAT 1660

RESULT 10
US-10-369-493-40734
; Sequence 40734, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfang
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40734
; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Caulobacter crescentus
US-10-369-493-40734

Query Match 12.7%; Score 225.6; DB 15; Length 1866;
Best Local Similarity 48.1%; Pred. No. 5.9e-60;
Matches 775; Conservative 0; Mismatches 814; Indels 21; Gaps 4;

QY 118 TGCGCTGATCTACGACGAGAAATGTCACAACTCCGCGCATCGACAGAGCTTG 177
DB 238 TGCGCGGCGCTACCTATGCGAGGCGCATCGCGGTTCAGAGGCGTCCGCACTGCTG 237
QY 178 CTTCCTTACGAGATTCGAGAGCGCTCGCTGCTTATGCTCTCTGAAATGACCTTGAA 237
DB 298 ATGATACGAAAGCTTACGAGCCCAAGACAGCGTATGATCTGTCGGGCAATTCGATCGAG 357

QY 238 CATTTGAGCTGGATTTGGGGCTATGATAGCGGCAATTCCTATTGCCGGTGTCTCT 297
DB 358 CAGGCTGATGAGCTGCGGCGCTTATACGCGGCGCTGCGGCGCGCATAGCCCG 417
QY 298 GCTTATTCATGCTGTGCGCAAGATTTGGCGAAGCTGCGTACATCTGAGTCTTCTGCA 357
DB 418 GCCTACAGCTGATCTGACCGACGACGCAAGTTCAGAGCATCTGCTTCAAAAGTGGCG 477
QY 358 CCGGAGCTGCTTTTGTGCGCATGACGACCTTTCAGAGGCGCAATTCAGATTCG 417
DB 478 CCGGCTGCTCTTTCGCGCAAGCGGAGCATGTTCCCGGGGCTTCGCGCACCTTAAG 537
QY 418 CCGGAGCATGCGCGCAATCTTCACTGAGGGGAATTTGCGCGGCGCGCAAGTACT 477
DB 538 GCGCTGAGCCGAGCTCTCTGTATATCACCGCATGCGGCGGCGGAGCGAGATCAC 597
QY 478 TTTGACAGCTGTGAGACAGCTGTGAGATTTGAGCAGATTAATGCTTTGCGGCAACT 537
DB 598 TTGCGCGAGGTTCGCGCGGACCAAGCGGCGCGCTGCGCGCGCGGAGATGCGGTG 657
QY 538 GCGCGGATACGATTCGCAAGTTCTTGTCTTCTGCTTCTGCTTCAAACTGCTTAAGCG 597
DB 658 GCGCGCGGAGGCTGCGCAAGTATCTTCACTCGGCTCAACGAGGCTTCGCGCAAGCGC 717
QY 598 GTGCGCATCTAGAGCAATGCTCT---GGGCAATAGAGATGCTTTCGCAAACTTC 654
DB 718 GTGCGCGAGCCAGCGCATGATGAGCGGCTATCGCGGCGAGAGGCGCTGCGCAC 777
QY 655 CCGGTTTTTGTGAGAGAGCGCGGCTGTGAGATGTTGCGTGTGAGAACACACTTC 714
DB 778 GAGGTTCGAGCAGACAGAGTTCGCCAGAGCTGGAATGATGCTGTGAGACCATCTCG 837
QY 715 GCGGCGACCAAAATCGGATGCTGTGTGTATCAACGCGCGACGTACTACCTTGACGAC 774
DB 838 GCGGCGCAATCGGCTTCAACGCGGTGATCTGAGGCGCGGCAACCTTCGACATCGACG 897
QY 775 GTTAAACCAACCGCCCAAGGTTTCGCGAGACGCTTCGCAACTGAGCGAAATCTCTCC 834
DB 898 GCGAGCGGCTGCGCGGATGTCGAGACCAAGATCAAGAACTTACGAGGTTCGCG 957
QY 835 ACTGCGACTCACTGTGCGGAAAGGCTGAGGAAATTAAGTGGTGCCTTGAAGCGAGAC 894
DB 958 ATGCTGTGCGGCTGCGCGGATGCTGCTTCTGATGCTGCGGAGGATGAGGAAGAC 1017
QY 895 AGTACCTGCGGAGAGCTTCTTCTGCTGCGATGAAGTGTCTTCTTTCGCGCGGCTGCG 954
DB 1018 CTGTGCTGCGGCTGCTTCTTCAAGAACTGCGCTACATAGGCTATGCGCGCGCAC 1077
QY 955 TTGTGCGAAGGATTCGGAATGCTTTGAGACCGGCTGCTGAAAGACATGCTGTGAGCG 1014
DB 1078 CTAGGCAATGACGCTATGAGCGCATCAGGCGCTGCGCGGAGACCGGCGCATGCG 1137
QY 1015 ATTGCAATGAGCGGCTGCGGATGCGGATGCGGAGAGTCTCTTCTGCACTTTTACAC 1074
DB 1138 ATTGCGTACCAAGATGACGCGGAGCGGAGGAGGAGGCTTCTATTAAGTGGCG 1194
QY 1075 GAGCGGCTGAGTGTGCTTACATTTGAGGCTGCGAGCGCTGCTGCGAGTCAAGCTC 1134
DB 1195 TGATACCGAGGAGGCTGCGGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 1254
QY 1135 GTTCGCGATGAGGAAATTTGAAAGGCGCTTTCATAGTTCGCAAGCATAGACGCTAC 1194
DB 1255 GCGCGGAGCGGCTGAGGATGAGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 1314
QY 1195 TGCGTGTCTCTGAAACAAATGCGCAAGGCTTTCAGAGGAGGCTTATCTGCTCGGT 1254
DB 1315 CAGAGAGCCCGAGGAGCGGCGGCGCTTTCAGAGGAGGCTTCTATTAAGTGGCG 1374
QY 1255 GATGCAATCAAAATGCGAGATCTGCGCATCTCAAGAAAGTGTATGTTTGAAGCTGCA 1314
DB 1375 GAGCGGCGGCTGCTGCTGATCGGAGGAGCCGCGCAAGGAGGCTGCTTTCAGAGGCGG 1434

QY 1315 ATTGCTGAAGACTTCAAGCTGTCTCAGAGGGATATTGTGACGGCTGGCCATTGCCAAG 1374
DB 1435 GTGACCGAGAGACTTCAAGCTCGACAGCGGTAAGTGGGTGAGCGTGGGGTGTGCGCC 1494
QY 1375 CGGCGGCTTCTGGAAGCGCGCTTTACGTCTTGTGAGAGTGTGTGTCTCTGATCGT 1434
DB 1495 GATCTGTGTGCGCGCTGACAGCCCTTCAATCAAGAGCGGTGTATCAACCGCCAGAGAC 1551
QY 1435 GAATGCTTGAATTTGCTGCTGTTTCCGCGCTTCTGCACTGCGGCTTGTGGAGCTTA 1494
DB 1552 -AAGGCTTTCATGGGCGGCAATGCTCTGGCCCTCGCGCGGCGGCGGCTTATGAGC 1610
QY 1495 GGAAGAGAGCGCTGGAAGCGCGGAGTGTCTTCCAGTGAAGCGGTGCGGCTGTGCT 1554
DB 1611 CGATCTGCGGCGCGGAGCGCGCTG-----GAGAGCTGTGTCGGATCTCAAG 1659
QY 1555 GACTGCTCAAGAGACTCAATCGAAGAGAACTGGCAATGCCATGCCATCATGTGGTA 1614
DB 1660 GACCGTCTGCGCGCTTTAAAGCGCGCGCGGCGGCTTTCGCGCGGCTGTGCGCGGCTT 1719
QY 1615 GGGCTCTCGATAGCGCGCGCTGATTTGATTAAGGCGAGGTCACTGACAAGGCTCGATC 1674
DB 1720 ACATCTCTGACCGAGCGCGCTGATCGAGCGCGGCGGAGATCAACGACAAGGCTTATC 1779
QY 1675 AACGAGCGCGCTGTTTTCGAATGCGCGGCGGAGAAAGTTGATGCGCTGTA 1724
DB 1780 AACGAGCGCTGCTGTGAGAGCGCGCGGATGATGTGAGGCGCTGTA 1829

RESULT 11
US-10-716-803-5
Sequence 5, Application US/10716803
Publication No. US20040229236A1

GENERAL INFORMATION:

APPLICANT: Blanc, Veronique
Blanche, Francis
Crouzet, Joel
Jacques, Nathalie
Lacroix, Patricia
Thibaut, Denis
Zagorec, Monique
Debusche, Laurent
De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved in The
Biosynthesis Of Streptogramins, Nucleotide Sequences
Coding For These Polypeptides And Their Use

NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farbow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.

COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/716,803
FILING DATE: 20-Nov-2003

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/635,359
FILING DATE: 09-AUG-2000
APPLICATION NUMBER: US 09/231,818
FILING DATE: 15-JAN-1999
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992

ATTORNEY/AGENT INFORMATION:
NAME: Arriago, Salvatore J.
REGISTRATION NUMBER: 46,063
REFERENCE/DOCKET NUMBER: 03806,0054-04000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4000
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1879 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaeapiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 110..1858
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-716-803-5

Query Match 3.2%; Score 57.4; DB 18; Length 1879;
Best Local Similarity 56.7%; Pred. No. 5.5e-07;
Matches 106; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1071 CACCGAGCGCTGTGATGGCTGTGATGTTGAGGCTGCCAGCGCTGCGAGGTCAA 1130
DB 1186 CACCGAGCGCGCTGTGATGTTGAGGCTGCCAGCGCTGCGAGGTCAA 1245
QY 1131 GCTGTTCCGCTGATGAGGAAATTTGAGGCGGTTTTCATGTCCGACGTATGAGCG 1190
DB 1246 GCCCGTCCCGCGGAGAACCGGTGATGCTACCGCGGCGCTTACACGCTGCGCG 1305
QY 1191 CTACTGCGCTGCTCTGAAACAAATTCGCAAGCGTTTCCAGAGAGGCTATTACTGCTC 1250
DB 1306 CTACTACCGGCGCGCGAGACAAACCGCGCGGTTTCCAGAGAGGCTTTCACGCGAG 1365
QY 1251 CGGTGAT 1257
DB 1366 CGCGAT 1372

RESULT 12
US-10-338-110-119

Sequence 119, Application US/10338110
Publication No. US20040023254A1
GENERAL INFORMATION:

APPLICANT: Fuhrmann, Jeffrey J.
TITLE OF INVENTION: A Method To Assess Quorum Sensing Potential Of Microbial
TITLE OF INVENTION: Communities

FILE REFERENCE: HER-0056
CURRENT APPLICATION NUMBER: US/10/338,110

CURRENT FILING DATE: 2003-01-07
NUMBER OF SEQ ID NOS: 123

SOFTWARE: PatentIn version 3.2
SEQ ID NO 119

LENGTH: 536
TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus Sequence

FEATURE:
NAME/KEY: misc_feature

LOCATION: (7)..(7)
OTHER INFORMATION: n is a, c, g, or t

FEATURE:
NAME/KEY: misc_feature

LOCATION: (21)..(21)
OTHER INFORMATION: n is a, c, g, or t


```
FEATURE:
NAME/KEY: misc feature
LOCATION: (33)..(33)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (69)..(69)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (87)..(87)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (213)..(213)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (226)..(226)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (261)..(261)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (273)..(273)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (347)..(347)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (357)..(357)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (398)..(398)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (419)..(419)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (509)..(509)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (517)..(517)
OTHER INFORMATION: n is a, c, g, or t
US-10-338-110-119

Query Match
Best Local Similarity 11.7%; Score 46.8; DB 16; Length 536;
Matches 54; Conservative 220; Mismatches 187; Indels 0; Gaps 0;

QY 382 GCAGCACTTTCCGCGCGCATTTCTGCGCGACGATTCGCGCGCATTTCC 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 73 SVRSYBVCWYSKMNWVVGDMMSBKVGARYKBSGAYCARTWCGAYWVYMSAHRGSK 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 442 ACTCGAGGCGAATTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 501
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 TACVTBTSKSHKVRWVSVVVBDAHATSKYGGHGYGSGSGCTGHTDCBACACV 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 502 GGTGGGATTGAGGAGATATGCTTTCGCGCGCACTTCGCGCGCATTTGCAAGTTC 561
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 193 BSSCBMAYCTGCTBRARKVNSTYTTTRSSBMVYSKCBVSRVYVMTTRCCSMVAT 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 562 TTGTTCACCTTCTGCGCTTACCAACTGCTTAAGCGCGTGCAGTACTCAGCGAATGCTC 621
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 253 CVMKSRNTNNGGABHTKWCNCGMTWCRGRCMCRGHSSSAMVYKGGCGRTGKVKRTY 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 622 TGCGCAATCAGAGATGCTTCTGCAACTTCCCGGTTTGTGTAAGACCGCGGCTG 681
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 313 TSKSDVMBCTSVAPMSGSDSYWBSGNGCVVRKBVSNTRTSRYSYRYSW 372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 682 CTGTGACTGTTGCGCGTGAACCAACACTTTCGCGCGCGCGCGCGCATTCGCGATGCTG 741
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 HCRBSRSSATGAVCGSYMTWCVDMKMSAMCGBRVTMDSHSSANCGVSYGGBVCGS 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 742 TTGTACAGCGCGCGCGCATCTTACCTTGACGACGCTTAACCAACCGCGCGGTTGCGC 801
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 433 YBMSVWVHVRBRBRVDMHRTVTSSTGCGSTSYGGAYCVBADCRRMDSVYVMSVHY 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 802 GAGACGCTTCCGAATTTGAGCGGAATCTCCACGCGCGTA 842
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 493 KWSCKYRSSMHYCVNRHBBGVBNMVKYBVGVDBRYKV 533
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-10-369-493-42307
; Sequence 42307, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 42307
; LENGTH: 1566
; TYPE: DNA
; ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-42307

Query Match
Best Local Similarity 49.2%; Score 46.8; DB 15; Length 1566;
Matches 123; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 517 GATATGCTTTTCGCGCACTGCGCGCGATTCGCAATTCGCAATTCGCTTCTGCGC 576
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 439 GACGACACGGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 498
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 577 TCTACCAACTGCTTAAGCGCGTCCGACTACTCAGCGAATGCTTTCGCGCGCATCAGCAG 636
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 499 ACCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 558
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 637 ATGCTTCTGCAACTTTCGCGGTTTTTGTGTAAGACCGCGGCTGCTGTAAGTGGTTG 696
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 559 CAGTTCATGACACTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 618
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 697 CCGTGAACCACTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 756
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 619 CCGTGTTCACACTTACCGCGCATGACCGTGGCATGAACCGCGCGCTGTCGCGCGCGCG 678
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 757 ACGTACTACC 766
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 679 ACGTCTTACC 688
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-10-282-122A-30430
; Sequence 30430, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
```


[illegible]

Db	1322	GAAACATTCAAGCTGTCTCAGGGGGATTMTTGTCAGCGTTGGGCACTTGCAGCCAGCGGCGC	1380
QY	1381	GTTTCTGGAAGCGGCGCTCTTACGTCTCTGAGAGTAAGTGTGTGTCTCTGATTCGTGAATATGC	1440
Db	1381	GTTTCTGGAAGCGGCGCTCTTACGTCTCTGAGAGTAAGTGTGTGTCTCTGATTCGTGAATATGC	1440
QY	1441	CTTGAGATTGCTCGTGTTCCGCGCGTCTTCTGACATCGCGGTGCTGTCTCGGGGCTAGAAAA	1500
Db	1441	CTTGAGATTGCTCGTGTTCCGCGCGTCTTCTGACATCGCGGTGCTGTCTCGGGGCTAGAAAA	1500
QY	1501	GAGCGCTCGAGCGCGCAGAGTGCTTGCCAGTAGACCGGTTTCGGGCTTGAGTTGTCTGACTGG	1560
Db	1501	GAGCGCTCGAGCGCGCAGAGTGCTTGCCAGTAGACCGGTTTCGGGCTTGAGTTGTCTGACTGG	1560
QY	1561	CTCAAACGATCTCAATTCGAGAAAGCACTGGCAATCCGACGTGCGATCATGTGGGTATGGGCTC	1620
Db	1561	CTCAAACGATCTCAATTCGAGAAAGCACTGGCAATCCGACGTGCGATCATGTGGGTATGGGCTC	1620
QY	1621	CTCGATACGCGCGCGCTGATGTATTAAGGCGAGGTCACCTGACAAAGGGCTCGATCAACACAG	1680
Db	1621	CTCGATACGCGCGCGCTGATGTATTAAGGCGAGGTCACCTGACAAAGGGCTCGATCAACACAG	1680
QY	1681	CGCGCTGTTTTCGCAATGCGCGGTTCGGCAAAAGTTGATCGCGCTGATTCGTGTGAAGATCAA	1740
Db	1681	CGCGCTGTTTTCGCAATGCGCGGTTCGGCAAAAGTTGATCGCGCTGATTCGTGTGAAGATCAA	1740
QY	1741	TTCGATGCTGCGTGACGAGGCCACCACTGTGA	1770
Db	1741	TTCGATGCTGCGTGACGAGGCCACCACTGTGA	1770

```

RESULT 2
US-08-976-063E-1
Sequence 1, Application US/08976063E
Patent No. 6524831
GENERAL INFORMATION:
APPLICANT: Steimbuchel, Alexander
APPLICANT: Priefert, Horst
APPLICANT: Rabenhorst, Jürgen
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
AND VANILLYL ALDEHYDE, FERULIC ACID, VANILLIN AND
TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
FILE REFERENCE: Bayer-9998-CAO
CURRENT APPLICATION NUMBER: US/08/976, 063E
CURRENT FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 186 49 655.1 GERMANY
PRIOR FILING DATE: 1996-11-29
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 32679
TYPE: DNA
ORGANISM: Pseudomonas sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (3146)..(3997)
OTHER INFORMATION: gene = "ORF1"
US-08-976-063E-1

```

	Query Match	Best Local Similarity	Score 1770;	DB 4;	Length 32679;
	Matches 1770;	Conservative	0;	Mismatches	Indels
			0;	Gaps	0;
Oy	1	ATGGCTTCTCTCGAGCGCTTCTTCCTTCCCGGGTCGAATTCCTTAGCGCTTCGAGCAT	60		
Db	21388	ATGGCTTCTCTCGAGCGCTTCTTCCTTCCCGGGTCGAATTCCTTAGCGCTTCGAGCAT	2134		
Oy	61	TGGGCTAAGACCGGTCAGAAACAACCTGGCTGTGTGCCAGGCGGCAATTGAGGAATG	120		
Db	21348	TGGGCTAAGACCGGTCAGAAACAACCTGGCTGTGTGCCAGGCGGCAATTGAGGAATG	2140		
Oy	121	CGTCGATCAGCTACGCGGAATGTTCCACAACGTCGCGCCATCGCACAAGCTTGCTT	180		

Db 21408 CGGTGATCAGCTACGGGAAATGTTCACAAAGCTCCGGCCATCGACAGAGCTTCTT 21467
Qy 181 CTTTACGAGATATCGGACAGAGCGTCCGCTGTTATGCTCTGTGAAATGACCTGGAACAT 240
Db 21468 CTTTACGAGATATCGGACAGAGCGTCCGCTGTTATGCTCTGTGAAATGACCTGGAACAT 21527
Qy 241 CTTTACGAGATATCGGACAGAGCGTCCGCTGTTATGCTCTGTGAAATGACCTGGAACAT 300
Db 21528 CTTTACGAGATATCGGACAGAGCGTCCGCTGTTATGCTCTGTGAAATGACCTGGAACAT 21587
Qy 301 TATTCACTGCTGTGCAAGATTTGGCGAAGTGGCTGACATCTGAGTCTTTCGCAACCG 360
Db 21588 TATTCACTGCTGTGCAAGATTTGGCGAAGTGGCTGACATCTGAGTCTTTCGCAACCG 21647
Qy 361 GAACTGCTTTTGTGCTGCGATGACAGACTTTTCAGGCGCAATTGAGACCATTTCTGCG 420
Db 21648 GAACTGCTTTTGTGCTGCGATGACAGACTTTTCAGGCGCAATTGAGACCATTTCTGCG 21707
Qy 421 GAGAGCGTGCCTGCAATCTTCACTGAGGCGAATTGGCCGGGGCGGCGAGGTGATTTT 480
Db 21708 GAGAGCGTGCCTGCAATCTTCACTGAGGCGAATTGGCCGGGGCGGCGAGGTGATTTT 21767
Qy 481 GAGAGCGTGCCTGCAATCTTCACTGAGGCGAATTGGCCGGGGCGGCGAGGTGATTTT 540
Db 21768 GAGAGCGTGCCTGCAATCTTCACTGAGGCGAATTGGCCGGGGCGGCGAGGTGATTTT 21827
Qy 541 CCGGATTCGATTTGCGCAAGTTCTTGTTCATCTTCTGCTTACCAACTGCTTAAAGCGGTG 600
Db 21828 CCGGATTCGATTTGCGCAAGTTCTTGTTCATCTTCTGCTTACCAACTGCTTAAAGCGGTG 21887
Qy 601 CCGGATTCGATTTGCGCAAGTTCTTGTTCATCTTCTGCTTACCAACTGCTTAAAGCGGTG 660
Db 21888 CCGGATTCGATTTGCGCAAGTTCTTGTTCATCTTCTGCTTACCAACTGCTTAAAGCGGTG 21947
Qy 661 TTTGCTGAAAGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 21948 TTTGCTGAAAGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22007
Qy 721 AGCCACAACATCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 22008 AGCCACAACATCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22067
Qy 781 CCAACGCGCCAAAGGCTTCCGCAAGCTTCCGCAAGCTTCCGCAAGCTTCCGCAAGCTTCC 840
Db 22068 CCAACGCGCCAAAGGCTTCCGCAAGCTTCCGCAAGCTTCCGCAAGCTTCCGCAAGCTTCC 22127
Qy 841 TACTCACTGTGCGAAAGCTGCGGAGAAATTAGTGGTGCCTTTCGAGGAGAGAGTACC 900
Db 22128 TACTCACTGTGCGAAAGCTGCGGAGAAATTAGTGGTGCCTTTCGAGGAGAGAGTACC 22187
Qy 901 CTGCGCGCAAGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 22188 CTGCGCGCAAGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22247
Qy 961 CAAAGGATCTGGAATCGTTTGGACCGGGTGCCTGAACAAGCATGCTGCTGCTGCTGCTGCT 1020
Db 22248 CAAAGGATCTGGAATCGTTTGGACCGGGTGCCTGAACAAGCATGCTGCTGCTGCTGCTGCT 22307
Qy 1021 ATGATGCGGGATCTGCGCATGACGAGACTGCTCTTCTGCACTTTTACACCGGACCG 1080
Db 22308 ATGATGCGGGATCTGCGCATGACGAGACTGCTCTTCTGCACTTTTACACCGGACCG 22367
Qy 1081 CTGCTGATGCTGTTACATTTGGGCTGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 22368 CTGCTGATGCTGTTACATTTGGGCTGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 22427
Qy 1141 GTGATGAGAAATTGGAAGGCGTTTCCATGCTGCGACGCGCAAGAGGCTTACCTGAGGT 1200
Db 22428 GTGATGAGAAATTGGAAGGCGTTTCCATGCTGCGACGCGCAAGAGGCTTACCTGAGGT 22487
Qy 1201 GCTCTGTAACAAATGCGCAAGCTTTCAGACGAGAGAGCTTATCTGCTCGGTGATGCC 1260
Db 22488 GCTCTGTAACAAATGCGCAAGCTTTCAGACGAGAGAGCTTATCTGCTCGGTGATGCC 22547

Qy 1261 ATCAAAATGCGAGATCCCTGCGATCCCTCAGAAAGTCTGATGTTTGAAGGCTCGAATTGCT 1320
Db 22548 ATCAAAATGCGAGATCCCTGCGATCCCTCAGAAAGTCTGATGTTTGAAGGCTCGAATTGCT 22607
Qy 1321 GAAGACTTCAAGCTGCTCTCAGAGGAGTATTTGTGACGCTTGGGCGCATTCGACGCGGCG 1380
Db 22608 GAAGACTTCAAGCTGCTCTCAGAGGAGTATTTGTGACGCTTGGGCGCATTCGACGCGGCG 22667
Qy 1381 GTTCTGAAAGCGGCTCTTACGCTCTGAGACGTAAGTGTCTGCTCTGATCTGAAATGC 1440
Db 22668 GTTCTGAAAGCGGCTCTTACGCTCTGAGACGTAAGTGTCTGCTCTGATCTGAAATGC 22727
Qy 1441 CTTGGAATGCTCGTGTTCGCGCTTCTGCACTGCGCTGCTTGTGCGGCGGATAGAAA 1500
Db 22728 CTTGGAATGCTCGTGTTCGCGCTTCTGCACTGCGCTGCTTGTGCGGCGGATAGAAA 22787
Qy 1501 GAGGCGTCCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Db 22788 GAGGCGTCCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22847
Qy 1561 CTCAAAGCACTCAATCGAAGAGCACTGCGCAATGCGCATGCTGATGAGGCTC 1620
Db 22848 CTCAAAGCACTCAATCGAAGAGCACTGCGCAATGCGCATGCTGATGAGGCTC 22907
Qy 1621 CTGATTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Db 22908 CTGATTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22967
Qy 1681 CGCGCTGTTTGGCAATGCGCGGCGCGCAAGTTGATGCTGCTGCTGCTGCTGCTGCT 1740
Db 22968 CGCGCTGTTTGGCAATGCGCGGCGCGCAAGTTGATGCTGCTGCTGCTGCTGCTGCT 23027
Qy 1741 TCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1770
Db 23028 TCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 23057

RESULT 3
US-09-328-352-2393
; Sequence 2393, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2393
; LENGTH: 1902
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2393

Query Match 25.2%; Score 445.4; DB 4; Length 1902;
Best Local Similarity 55.2%; Pred. No. 1.5e-118;
Matches 932; Conservative 0; Mismatches 746; Indels 9; Gaps 3;

Qy 46 GAGGCTGAGCAATGAGGCTGATGAGCCGCTGCAAGCAAACTGCGTGTGCGCAGGCG 105
Db 166 GATGCTTAAATTTGATTTTGCACAAACCAACCCGACATATTTTTCAGCAAAAGCAAT 225
Qy 106 GCAATGAGGAAATGCGCTGCTATCAGCTACGCGGAAATGTTTCAACAAGTCCGCGCATC 165
Db 226 GCTCAAGGCAATGAGGCTGCAATGAGTTATGCAAGAGTTCTTACAAACGCGATGACATTT 285
Qy 166 GCACAGAGCTTGTCTCTTACGAGACTATGCGAGAGAGTCCGCTGCTTATGCTCTGCA 225
Db 286 GCTCAAGCTTGTATGCTGCTTAAATTTTAAAGCAAGAAAGACTTGTGCTATTTAAGTGT 345
Qy 226 AATGACTGGAACATTTTCACTGCTGCTTGTGGGCTATGATGCGGAGCATTCCTATTCG 285

Db 346 AATGATCTCGAATTTAACAAGTCTATGCTGCGCAATGCTGGCAGCGCTTTCTCG 405
Qy 286 CCGGTCTCTCTCTATTTACAGTCTGTCGCAAGATTTGGCGAAGCTGCACTGTA 345
Db 406 GCTATTTTCCCTCTACTCTTGATTTCTCAAGAATTGGCAAACTGCAAACTGTTT 465
Qy 346 GGTCTTCTGCAACCGGAGTGTCTTGTCTGCGCATGCAACCTTTCCAGCGCAATT 405
Db 466 GAGTGTCTACACTGCTGATGTTTATGCGAGGATGGAACAAGCTTTTGGCAAGGAT 525
Qy 406 GAAACATTTTGGCGGACGAGCGTCCGCAATCTTCACTGAGCGCAATTGGCCGGCG 465
Db 526 CAGGCAATGATTAAGCTGATTAATGAAGTGAACCAATTAAGAAATAGGCGCATCAG 585
Qy 466 CGCAGGTGATTTTGAACGCTGCTGAGCAGCTGCTGGGATTTGAGCGAATTAATGCC 525
Db 586 ATTTGCACTCTTTTCAATGCTGTTAATGATTAACCAAGTTTCAATGTTCAAGGTTT 642
Qy 526 TTTCGCGCACTGCCCCGATTAAGATTGCAAGTTCTTGTTCATCTTCTGCTTACCAA 585
Db 643 TATCAAACTCTTATGAATAACAGATTGCAAAATTTCTGTTTCACTGAGTTCACTAA 702
Qy 586 CTGCTTAAGCGGTGCGGACTACTCAGCGAATGCTGCGCAATCAGCAGATGCTTCTG 645
Db 703 TTACTTAAGCTGTACGACACACATTTAATGTTGTGTTAATCAGCAAAATGTTATTG 762
Qy 646 CAAACTTTCCGGTTTTTGGTGAAGACCGCGGTGCTGAGACTGTTGGCTGGCAAC 705
Db 763 CAGACTTTTCCCTAGTTTGAAGAAACAGCGCTGTCTTACTCGACTGCGTGTCTTGGCAC 822
Qy 706 CACACTTTCGGCGGACCCCAACCATCGCATGCTGTTTGAACAAGCGGCGACGTAATAC 765
Db 823 CACACATTTTGGCGGAGTCACAATGTCGATCGACTATTAACCGCGGTACGATTTAC 882
Qy 766 CTGGAAGCGGTAAACCAACCGCCCAAGGTTTGGCGAAGCGCTTGGCACTTGAAGAA 825
Db 883 ATTTGATGATGCAAAACCGCTTGGCAGGAAATTTGACAAACTTATCTGTAATCAAGAA 942
Qy 826 ATCTCTCCCACTGTAATCTCACTGTCGCAAAAGGCTGGAGGAATTAAGGCTGACCTT 885
Db 943 ATTTCTCCAACTGTTTATTTAAATGTCGCAAAAGTTGGAGAACTACCGAAGCTTA 1002
Qy 886 GAGCGAGACGTAACCTGCGCGAAGCGTTTCTGCTGCAATGAAGCTGTTCTTTCGCG 945
Db 1003 GAAAGATGAAGATTAAGAGCGCTTTTGGCAAACTTAAATTTATCTTTCGCC 1062
Qy 946 GCGGCTGGTTCGCAAGGATCTGGGATCTGTTGACCGGCTGCTGAACGCACTGT 1005
Db 1063 GGTGCTGCACTTTCAGAAAGGCGCTGGAACAGACTGATTAATTTGCTCAGCAACTTGC 1122
Qy 1006 GGTGAGCGCACTTTCGATGATGCGGCTGCTGGGATGACGAGACTGCTCTTCTGCACT 1065
Db 1123 GGAAGAAAATTCGCACTTATGAGCGGATTTGGGCAATGACGAACTGCTCTTGTGCT 1182
Qy 1066 TTATACACCGGACGCTGCTGATGCTGATTAATGAGGCTGCAAGCGCTGCTGGAG 1125
Db 1183 TTATACACTGGCCAGCGCTAATGCTGCTTATTGTTTACCTGCTCGGGATGGAA 1242
Qy 1126 GTTAAGCTGTTCCGCTGATGAGAAATTTGAAGGCGCTTTCATGCTCGCAGCTCATG 1185
Db 1243 ATTAACTAATTCATGCTGATGCAAACTTGAAGTTTTCGCTGTCGCAAAACGCTCATG 1302
Qy 1186 AGCGGCTACTGGCG---TGCTCTGAAACAAATGCCCCAAGCTTTCGACGAGAGGCTAT 1242
Db 1303 AAAGGCTATTGGCGCTTAAAGCGGACCAACAAATGTAATTTTGTATGATGAAGGCTTT 1362
Qy 1243 TACTGCTCGGTATGCAATTAATGGCAATCTGCGCATCTCGAAGGCTGTGATG 1302
Db 1363 TTCCATATCAGGCGATGCGTTGCTTATGCTGATGTAATCAACCAAAAGGCTAATG 1422
Qy 1303 TTGAGCGGTGATGCTGAAGACTTAAGCTGCTCAGGGGATTTTGTCAAGCTTGGG 1362

Db 1423 TACAGCAGCAATTECCAGAAACTTTAACTCAATACAGGCACTTTGTGAATGTCGC 1482
Qy 1363 CCATTCGCAAGCGGCGGCTTCTGAAGCGGCTTACGTCCTGGACGTAGTGTGCT 1422
Db 1483 ACATCAGCAACAAAGCTTATTCAGGTAATTTACTATCCAGATGTTTGTATACT 1542
Qy 1423 GCTCTGATGTAATGCTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1482
Db 1543 GGTTCAAACCTGAATGCTATGCTGTTTCTGATTTTTCAAAATTAACCTGCTGTCAA 1602
Qy 1483 TTCTCGGCTTGAAGAAAGGCG---TCGAGCGGAGGTGCTTCCAGTACCGCGTT 1539
Db 1603 TATCAGGTCTTAAGCTGGCGCAATTTCTGACGAGAAATCTGCAACACCTTAAGTC 1662
Qy 1540 CCGGCTGCTTGTCTGACTGCTCAACAGCTCAATCGAAGCACTGGCAATGCAAGT 1599
Db 1663 CAACATGCTTCCGCAATTTTAAAGACCTTAAATTAAGATGCGACTGCGACTCAAT 1722
Qy 1600 CGCATATGCTGATGAGGCTCTCTGATACCGCGCTGCTGATGATTAAGGCGAGTCACT 1659
Db 1723 ACAGTCTCAATGCTTATTTAATGAACGAGCACTCAGTTAGATGCGCGCAAGTAC 1782
Qy 1660 GAAAGGCTGATCAACAGCGCGCTGTTTGCATGAGCGGTGCGGCAAGTTGATGCG 1719
Db 1783 GATTAAGGCAACTCAATCAAGCAAGTATTAACCAAGCTGCTGCTTTAATGATGAG 1842
Qy 1720 CTGTATC 1726
Db 1843 CTTTATC 1849

RESULT 4
US-08-403-852D-5
; Sequence 5, Application US/08403852D
; Patent No. 5891695
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Jacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crecy-Jagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,852D
; FILING DATE: 10-MAY-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR 93/00923
; FILING DATE: 25-SEP-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: FR 92/11441
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1879 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ANTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 110..1858
US-08-403-852D-5

Query Match 3.2%; Score 57.4; DB 2; Length 1879;
Best Local Similarity 56.7%; Pred. No. 1.9e-06;
Matches 106; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1071 CACCGACCGCTGTCATGCTGTTACATTGGGCTGCCAGCGCTGCTGCGAGTCAA 1130
DB 1186 CACCCAGGCGCGCGGCTGTCGAGCGGCGAGATCGGCTGCCAGCCCGACGCGCG 1245
QY 1131 GCTCGTCCGCTGATGGAATTTGAAGGCGCTTTCATGTCGCGACGTCATGACGG 1190
DB 1246 GCCCGTCCCGCGGCTGAGACCGGTGAATGCTCACCCTGCGGCGGCTTCAACGCTGCGCG 1305
QY 1191 CTACTGCGCTGCTCTCTGAACAAATGCCCAAGCGTTTCGAGAGAGAGGCTATTACTGCTC 1250
DB 1306 CTACTACCGGCGCGCGGCGAGCAACGCCCGCGCTTCAACGAGAGCGCTTCTACCGCAG 1365
QY 1251 CGGTGAT 1257
DB 1366 CGGCGAT 1372

RESULT 5
US-08-510-646B-5
Sequence 5, Application US/08510646B
Patent No. 6077699
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crey-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flanagan, Henderson, Parabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,646B

FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1879 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ANTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 110..1858
US-08-510-646B-5

Query Match 3.2%; Score 57.4; DB 3; Length 1879;
Best Local Similarity 56.7%; Pred. No. 1.9e-06;
Matches 106; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1071 CACCGACCGCTGTCATGCTGTTACATTGGGCTGCCAGCGCTGCTGCGAGTCAA 1130
DB 1186 CACCCAGGCGCGCGGCTGTCGAGCGGCGAGATCGGCTGCCAGCCCGACGCGCG 1245
QY 1131 GCTCGTCCGCTGATGGAATTTGAAGGCGCTTTCATGTCGCGACGTCATGACGG 1190
DB 1246 GCCCGTCCCGCGGCTGAGACCGGTGAATGCTCACCCTGCGGCGGCTTCAACGCTGCGCG 1305
QY 1191 CTACTGCGCTGCTCTCTGAACAAATGCCCAAGCGTTTCGAGAGAGAGGCTATTACTGCTC 1250
DB 1306 CTACTACCGGCGCGCGGCGAGCAACGCCCGCGCTTCAACGAGAGCGCTTCTACCGCAG 1365
QY 1251 CGGTGAT 1257
DB 1366 CGGCGAT 1372

RESULT 6
US-09-231-818-5
Sequence 5, Application US/09231818
Patent No. 6171846
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crey-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
3 STREET: 1300 I Street, N.W., Suite 700
4 CITY: Washington
5 STATE: D.C.
6 COUNTRY: USA
7 ZIP: 20005-3315
8
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: Patentin Release #1.0, Version #1.30
14
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/09/231,818
17 FILING DATE:
18 PRIORITY APPLICATION DATA:
19 APPLICATION NUMBER: US/08/403,852
20 FILING DATE: 10-MAY-1995
21 APPLICATION NUMBER: PCT/FR 93/00923
22 FILING DATE: 25-SEP-1993
23 PRIORITY APPLICATION DATA:
24 APPLICATION NUMBER: FR 92/11441
25 FILING DATE: 25-SEP-1992
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Meyers, Kenneth J.
28 REGISTRATION NUMBER: 25,146
29 REFERENCE/DOCKET NUMBER: 03806.0054-00000
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: (202) 408-4000
32 TELEFAX: (202) 408-4400
33 INFORMATION FOR SEQ ID NO: 5:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 1879 base pairs
36 TYPE: nucleic acid
37 STRANDEDNESS: double
38 TOPOLOGY: linear
39 MOLECULE TYPE: cDNA
40 HYPOTHEICAL: NO
41 ANTI-SENSE: NO
42 ORIGINAL SOURCE:
43 ORGANISM: S.pristinaespiralis
44 FEATURE:
45 NAME/KEY: CDS
46 LOCATION: 110..1858
47
48 US-09-231-818-5
49
50 Query Match 3.2%; Score 57.4; DB 3; Length 1879;
51 Best Local Similarity 56.7%; Pred. No. 1.9e-06;
52 Matches 106; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

Db 1306 CTACTACCGGGCCCCGAGACACACGCCCGGCTTACCGAGACGGCTTCTACCGCAG 1365
|||
Qy 1251 CGGTGAT 1257
|||
Db 1366 CGGCGAT 1372
|||

RESULT 8

US-09-252-991A-4762/c
; Sequence 4762, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4762
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4762

Query Match 2.5%; Score 44.8; DB 4; Length 891;
Best Local Similarity 52.7%; Pred. No. 0.0057;

Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 1081 CTGTGATGCTGTGATTACATTTGGGCTGCCAGCGCTGCGAGTCAAGCTCTGTCG 1140
|||
Db 717 CGGCCCTGTGCGCGAGACGAAAGTGGGTGATCACCAGGCGCGAGTGCAG 658
|||
Qy 1141 GTGATGGGAAATTGGAAGGCGCTTTCATGCTCCGACGTCAAGCGGCTACTGCGCT 1200
|||
Db 657 CCGGGGAGGTGCGCGAACTGACCGTGGCGCTCCCTACACATCCCGGCTACTACCGC 598
|||
Qy 1201 GCTCTGAACAATATGCCAAGCGCTTTCGACGAGGAGGCTATTACTCTCCGATGCC 1260
|||
Db 597 CTGCGGACACACACGCAAGGCTTTCAGCGGAGCGCTTTCACCGACCGGACCGG 538
|||
Qy 1261 ATCA 1264
|||
Db 537 GTCA 534
|||

RESULT 9

US-09-252-991A-4723
; Sequence 4723, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4723
; LENGTH: 1554
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4723

Query Match 2.5%; Score 44.8; DB 4; Length 1554;
Best Local Similarity 52.7%; Pred. No. 0.0074;
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 1081 CTGTGATGCTGTGATTACATTTGGGCTGCCAGCGCTGCGAGTCAAGCTCTGTCG 1140
|||
Db 251 CGGCCCTGTGCGCGAGACGAAAGTGGGTGATCACCAGGCGCGAGTGCAG 310
|||
Qy 1141 GTGATGGGAAATTGGAAGGCGCTTTCATGCTCCGACGTCAAGCGGCTACTGCGCT 1200
|||
Db 311 CCGGGGAGGTGCGCGAACTGACCGTGGCGCTCCCTACACATCCCGGCTACTACCGC 370
|||
Qy 1201 GCTCTGAACAATATGCCAAGCGCTTTCGACGAGGAGGCTATTACTCTCCGATGCC 1260
|||
Db 371 CTGCGGACACACACGCAAGGCTTTCAGCGGAGCGCTTTCACCGACCGGACCGG 430
|||
Qy 1261 ATCA 1264
|||
Db 431 GTCA 434
|||

RESULT 10

US-09-252-991A-4682
; Sequence 4682, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4682
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4682

Query Match 2.5%; Score 44.8; DB 4; Length 1752;
Best Local Similarity 52.7%; Pred. No. 0.0078;
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 1081 CTGTGATGCTGTGATTACATTTGGGCTGCCAGCGCTGCGAGTCAAGCTCTGTCG 1140
|||
Db 1201 CGGCCCTGTGCGCGAGACGAAAGTGGGTGATCACCAGGCGCGAGTGCAG 1260
|||
Qy 1141 GTGATGGGAAATTGGAAGGCGCTTTCATGCTCCGACGTCAAGCGGCTACTGCGCT 1200
|||
Db 1261 CCGGGGAGGTGCGCGAACTGACCGTGGCGCTCCCTACACATCCCGGCTACTACCGC 1320
|||
Qy 1201 GCTCTGAACAATATGCCAAGCGCTTTCGACGAGGAGGCTATTACTCTCCGATGCC 1260
|||
Db 1321 CTGCGGACACACACGCAAGGCTTTCAGCGGAGCGCTTTCACCGACCGGACCGG 1380
|||
Qy 1261 ATCA 1264
|||
Db 1381 GTCA 1384
|||

RESULT 11

US-09-535-521-1
; Sequence 1, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
; APPLICANT: Weber, Eric R.
; APPLICANT: McCall, Catherine A.
; TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IGE RECEPTOR (CANINE CD23)
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF

```
FILE REFERENCE: AL-5
CURRENT APPLICATION NUMBER: US/09/535,521
CURRENT FILING DATE: 2000-03-24
EARLIER APPLICATION NUMBER: 60/125,913
EARLIER FILING DATE: 1999-03-24
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 2851
TYPE: DNA
ORGANISM: Canis familiaris
FEATURE:
NAME/KEY: CDS
LOCATION: (199)..(1077)
US-09-535-521-1
```

```
Query Match      2.4%; Score 41.6; DB 4; Length 2851;
Best Local Similarity 55.6%; Pred. No. 0.082;
Matches 80; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
```

```
QY 664 GGTGAAGAGCCGCGGCTGTGTGACTGTGCGCGTGAACACACCTTCGGGGGAGC 723
DB 976 GCGGAGGACTGCGTGTATGATGCGGGCTCGGGGAGTGAATACGCTTCTGCGGAGC 1035
QY 724 CACAACATCGGCATCGTGTGTGAACAAGCGGACAGTACTTGAAGACGGTAAACA 783
DB 1036 TCGCTGAGCGGTGGGTGTGTGACCGGCTGGCAGCGTGTGAGCCACCGGCGGCTTC 1095
QY 784 ACCGCCCAAGGGTTCCCGGAGAGC 807
DB 1096 GGCACCCCGGGGTTCCGTGGGAGT 1119
```

RESULT 12

```
US-09-535-521-3/C
Sequence 3, Application US/09535521
Patent No. 6410714
GENERAL INFORMATION:
APPLICANT: Weber, Eric R.
APPLICANT: McCall, Catherine A.
TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IGE RECEPTOR (CANINE CD23)
FILE REFERENCE: AL-5
CURRENT APPLICATION NUMBER: US/09/535,521
CURRENT FILING DATE: 2000-03-24
EARLIER APPLICATION NUMBER: 60/125,913
EARLIER FILING DATE: 1999-03-24
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 2851
TYPE: DNA
ORGANISM: Canis familiaris
US-09-535-521-3
```

```
Query Match      2.4%; Score 41.6; DB 4; Length 2851;
Best Local Similarity 55.6%; Pred. No. 0.082;
Matches 80; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
```

```
QY 664 GGTGAAGAGCCGCGGCTGTGTGACTGTGCGCGTGAACACACCTTCGGGGGAGC 723
DB 1876 GCGGAGGACTGCGTGTATGATGCGGGCTCGGGGAGTGAATACGCTTCTGCGGAGC 1817
QY 724 CACAACATCGGCATCGTGTGTGAACAAGCGGACAGTACTTGAAGACGGTAAACA 783
DB 1816 TCGCTGAGCGGTGGGTGTGTGACCGGCTGGCAGCGTGTGAGCCACCGGCGGCTTC 1757
QY 784 ACCGCCCAAGGGTTCCCGGAGAGC 807
DB 1756 GGCACCCCGGGGTTCCGTGGGAGT 1733
```

RESULT 13

```
US-08-804-227C-7/C
Sequence 7, Application US/08804227C
Patent No. 5876991
```

```
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuntz, Stuart A.
APPLICANT: Rostek, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSER: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
```

```
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
```

```
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
```

```
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7
```

```
Query Match      2.3%; Score 41; DB 2; Length 44377;
Best Local Similarity 52.7%; Pred. No. 0.43;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
```

```
QY 732 CGGATCGTGTGTGTACACGCGGACGTAATCACTTGAACAGCGTAAACACGCCCA 791
DB 1139 CGCGCTGTTGCGGACGCGACGCCACGACCAACGAGTACACGGGGTCCCGTCGCCCA 1080
QY 792 AGGTTCCGCGGAGACGCTTCGCACTTGAACGGAATCTTCCACTGCGTACTCTGT 851
DB 1079 GGGCGTGGGAGAGGGCTTCAGGACGACGCGCGCGCTTACCGGACATACGCGT 1020
QY 852 GCCAAGGCTGGAGAGATTAAGTGGTGCCCTTGAAGCAGATACC 900
DB 1019 TGGACGGGCGTCAAGGTGTGTGGCAGCGGCGTGGGTACACGGCGCC 971
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2005, 15:41:02 ; Search time 826 Seconds
(without alignments)
11248.759 Million cell updates/sec

Title: US-09-750-986D-29

Perfect score: 1770
Sequence: 1 atcgctctctcgagcgctc.....gtgacgagcgccacactgtga 1770

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues
Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_23Sep04:*
2: geneseqn1980s:*
3: geneseqn1980s:*
4: geneseqn2000s:*
5: geneseqn2001as:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	445.4	25.2	1902	9 ADA31106	ADA31106 DNA encod
2	404.6	22.9	1203	3 AAA39722	AAA39722 Pseudomon
3	397.2	22.4	2188	3 AAA39720	AAA39720 Pseudomon
4	396.8	22.4	2171	3 AAA39721	AAA39721 Pseudomon
5	73	4.1	1543	3 AAA39731	AAA39731 Pseudomon
6	73	4.1	2509	3 AAA39730	AAA39730 Pseudomon
7	73	4.1	2526	3 AAA39729	AAA39729 Pseudomon
8	57.4	3.2	1879	2 AA064205	AA064205 sRNA gene
9	53.2	3.0	2000	8 ADA71938	ADA71938 Rice gene
10	46.8	2.6	536	10 AD868842	AD868842 Minorily
11	46.6	2.6	5057	2 AA223757	AA223757 P. Fluore
12	46.4	2.6	1644	8 ACA42560	ACA42560 Prokaryot
13	45.6	2.6	2000	8 ADA71938	ADA71938 Rice gene
14	45.4	2.6	5451	5 AAF26319	AAF26319 Pseudomon
15	45.4	2.6	6522	8 ACA43663	ACA43663 Prokaryot
16	44.8	2.5	891	11 ABD06158	ABD06158 Pseudomon
17	44.8	2.5	1554	11 ABD06119	ABD06119 Pseudomon
18	44.4	2.5	1752	11 ABD06078	ABD06078 Pseudomon
19	44.4	2.5	1632	8 ACA51097	ACA51097 Prokaryot
20	43.4	2.5	342	8 ACA14026	ACA14026 Prokaryot
21	43.4	2.5	410	8 ACA15140	ACA15140 Prokaryot

22	43.4	2.5	1701	8 ACA18952	ACA18952 Prokaryot
23	43	2.4	10732	3 AAA10594	AAA10594 Gene enco
24	41.6	2.4	2851	6 ABN86559	ABN86559 Canine 1o
25	41.6	2.4	2851	6 ABN86560	ABN86560 Canine CD
26	41.4	2.3	1571	3 AAA39728	AAA39728 Pseudomon
27	41.4	2.3	2506	3 AAA39727	AAA39727 Pseudomon
28	41.4	2.3	2539	3 AAA39726	AAA39726 Pseudomon
29	41.2	2.3	9192	12 AD026464	AD026464 Brevibact
30	41	2.3	44377	2 AAT80414	AAT80414 Platenoli
31	41	2.3	44377	2 AAT78508	AAT78508 Platenoli
32	40.6	2.3	1754	4 AAF81366	AAF81366 Quorum se
33	40.6	2.3	13029	4 AAS51470	AAS51470 Pseudomon
34	40.6	2.3	13029	8 ACA19370	ACA19370 Prokaryot
35	40.2	2.3	1803	10 ACF71666	ACF71666 Phototrab
36	40.2	2.3	110000	10 ACF67367_48	Continuation (49 o
37	40.2	2.3	110000	10 ACF65387_0	ACF65387 Phototrab
38	40.2	2.3	349980	6 ABO81844	ABO81844 Bifidobac
39	39.6	2.2	1668	10 AB237954	AB237954 N. gonorr
40	38.8	2.2	1744	2 AA063900	AA063900 Acyl COA
41	38.6	2.2	2003	5 AAS85442	AAS85442 DNA encod
42	38.6	2.2	3286	6 ABS51519	ABS51519 Human CDN
43	38.6	2.2	4468	6 ABS51504	ABS51504 Human CDN
44	38.6	2.2	13855	10 ADE79055	ADE79055 Human pro
45	38.4	2.2	609	11 ABD08475	ABD08475 Pseudomon

ALIGNMENTS

RESULT 1
ID ADA31106 standard; DNA, 1902 BP.
XX
AC ADA31106;
XX
DT 20-NOV-2003 (first entry)
XX
DE DNA encoding Acinetobacter baumannii protein #2393.
XX
KW day; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
XX vaccine; plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
(GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
XX WPI, 2003-576092/54.
XX
DR P-PSDB; ADA35232.
XX
PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
XX for diagnosing a bacterial disease, as components of antibacterial
XX vaccines, as targets for antibacterial drugs, or as biocontrol agents for
XX plants.
XX
XX Example; SEQ ID NO 2393; 328bp; English.
XX
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
XX for diagnosing a bacterial disease, as components of antibacterial
XX vaccines, as targets for antibacterial drugs, to detect the presence of
XX A. baumannii and other Acinetobacter species in a sample, in screening
XX compounds for the ability to interfere with the A. baumannii life cycle
XX or to inhibit A. baumannii infection, and as biocontrol agents for

CC plants. The present sequence represents DNA encoding an A. baumannii
CC protein.

XX Sequence 1902 BP; 566 A; 420 C; 392 G; 524 T; 0 U; 0 Other;

Query Match 25.2%; Score 445.4; DB 9; Length 1902;

Best Local Similarity 55.2%; Pred. No. 1e-115;

Matches 932; Conservative 0; Mismatches 746; Indels 9; Gaps 3;

```

QY 46 GAGGCTCGAGATTGGGCTTAAGACCCCTCCAGAACCAACCTGCTGCGACAGGCG 105
DB 166 GATCGTTTAATTATTTTGCACAAACCAACCCGACCATATTTTGCAGCAAAACGCAAT 225
QY 106 GCAAATGGGGAATGCGCTCATAGCTACGCGGAATGTTCACAACAGTCCGCGCCATC 165
DB 226 GCTCAAGGCAATGGGTCACAACTGAGTTATGCAAGAGTTCTACAGCGCATGGCATT 285
QY 166 GCACAGAGCTTGCTTCTTACGAGCTATCGGCAAGAGCTCCGCTCTTATGCTCTGGA 225
DB 286 GCTCAAGCTTTGATGCTCGTAAATTTAAGCCAAAGAACCTGCTGATTTTAAAGTGT 345
QY 226 AATGACCTGGAAATCTTACGCTGGCATTTGGGGCTATGTATGCGGCAATTCCTATTCG 285
DB 346 AATGATCTGAACATTTTAACATGCTGTATGCTGCTGCTGCGAGCGCTGCTTCTCG 405
QY 286 CCGGTGCTCTCTTATTCATCTGCTGTGCGAAGATTTGGCGAAGCTGCTCATGCTGA 345
DB 406 GCTATTTCCCTGCTACTGCTGATTTCTCAAGACTTTGGCAAACTCAAACTATGTTT 465
QY 346 GGTCTTCTGCAACCGGCACTGTCTTTGCTGCGATGCAAGCACTTTCCAGCGCAATT 405
DB 466 GAAGTGCTCACACCTGGATGTTTATGTCAGGAGTGAACAACTTTTGCAAAGCGATT 525
QY 406 GAGACCATTTGCGCGGACGAGCTGCCGCAATCTTCACTGAGGCGCAATTTGGCGGCG 465
DB 526 CAGGCAATGATTAAGCTGTATTTGAAGTGTGACCAATTAAGAAATAGTGGCGATCAG 585
QY 466 CGGACGCTGATTTTGAACAGCTGCTGAGACAGCTGTGGGATTTGAGCGATTAATGCC 525
DB 586 ATCTGACGCTTTTCAATGCTGTATGATACACAGTTTCAATGTTCAAGAGTTT --- 642
QY 526 TTTGGCGCACTGCGCCCAATACGATTTGCCAATGTTCTTGTCACTTGTGCTCTACCAA 585
DB 643 TATCAAAACCTTATGAAACCAATGTGCAAAATTTCTGTTTATCATCAGGTTTCAACTAAA 702
QY 586 CTGCTAAGGCGGTGCGCACTACTACGCAAGTCTGCGCCCAATCAGCAGATGCTTCTG 645
DB 703 TTACTTAAGCTGTACCGACCAACATTTAATGTGTGTATTCAGCAATGTTATTTG 762
QY 646 CAAACTTTCCCGGTTTTTGTGAAGACCGCGGTGCTGTGATGCTGTGCGTGGAAC 705
DB 763 CAGACTTTCCCTGAGTTTGAAGAAACACGCTGTCTACTCGACTGCTGTGTCGAC 822
QY 706 CACACCTTGGCGGCGACCAACATGTGGCATGTGTGTGACAAACGCGGCACTACTAC 765
DB 823 CACACATTTGGCGGCGGTCAATGTGCGCATGCACTATTAACGCGGCGGATGATTTAC 882
QY 766 CTGACGACGCTTAACCAACCGCCCAAGGTTGCGCGACGCTTGCACATTTGAGCGAA 825
DB 883 ATTGATGATGCAAAACCGTGTGCAAGAAATTTGACGAATCTATGCTTCAAGAA 942
QY 826 ATCTCTCCACTGCTGTACTCTACTGTGCGAAAGGCTGGAGGAATTAATGTGGTCCCTT 885
DB 943 ATTTCTCAACTGTTTATTTAATGTGCAAAAGGTTGGGAAAGTCAACCGAAGCGTTA 1002
QY 886 GACGAGACAGTACCTCGCGGCAACGCTTCTGTGCTGCAATGAAGCTGTCTTCTTGGCG 945
DB 1003 GAAAAAGATGAAGATTAAGAGAGCGCTTTTGTCCAAAGTTAAATTTATTTCTTGGCC 1062
QY 946 GCGGCTGTGGTTGTGCAAGGAGATCTGGGATGTTTGGACCGGGTGTGCAACGCACTGT 1005
DB 1063 GGTGCTGCACTTTCAAGAGCGGCTGGACAGACTGATTAATTTGCTTCAGCAACTTGC 1122

```

```

QY 1006 GGTGAGCGCAATTCGATGATGGCGGTCTGGGCGATGACGAGACTGCTCTTCTGCACT 1065
DB 1123 GAGAAAAAATCCGCAATTATGAGCGGATTTGGGCGATGACCGAAACTGCTCCTCTTGCT 1182
QY 1066 TTTACCAACCGGACCGGCTGCAATGCTGTGTTACATTTGGGCTGCAAGCGCTGCTCGAG 1125
DB 1183 TTTTCAACTGGGCCCAAGCGTAAATGCTGCTTTATGTGTACCTGTCTCGGAGTGGAA 1242
QY 1126 GTCACCTCGTTCCGCTGATGGGAAATTTGAAGGCGGTTTTCCATGATTCGACGCTCATG 1185
DB 1243 ATTTAAGCTATTTGCGATGATGATGACCAACTTGAATTTGGCTGTGTGGCAACAGCTCATG 1302
QY 1186 AGCGGCTACTGCGG---TGCCTGTAACAATATGCCAAGCTTTCAGACGAGAGGCTAT 1242
DB 1303 AAGGCTATTTGGCGCTTAAAGGCGGACCAACAAGTACTATTTTGTATGATGAAGGCTTT 1362
QY 1243 TACTGCTCGGATGATGCAATTTGGAGATTCCTCGGATCCGATCTCAAGAAAGTGTGATG 1302
DB 1363 TTTCAATACAGGCGATCCGTTGCTTAAATGATGATCAATGATCAACAAAGGCTTAAATG 1422
QY 1303 TTTGACGCTGCAATTTGCTGAAGACTTCAAGCTGTCTCTCAGGGGATTTTGTACAGCTTGGG 1362
DB 1423 TACGACGAGCAAGATTTGCCAAGACTTTAACTCAATACAGGCACTTTTGTGATGTGCGG 1482
QY 1363 CCAATTCGCAACGCGGCGGCTTCTGAAAGCGGCTTTACGTCTTGAAGCTATGTGTTGCT 1422
DB 1483 ACACTACGCAACAAGATGCTTATTCAGGTTAAATTTACTATCCAGATGTTGTATTAAT 1542
QY 1423 GCTTCGATCGTAATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1482
DB 1543 GGTTCAAACCTGATGCTATTTGTTTCTGATTTTTCAAAATTTAAGCGTTGTGCTCAA 1602
QY 1483 TTTGCGGCGCTAGGAAAGAGCG--TCGACGCCGAGATGCTTTCAGTGAAGCGGATT 1539
DB 1603 TATGACAGCTTAAAGCTGAGCGCAATTTCTGACAGCAAGATTACTGCAACACCTTAAATG 1662
QY 1540 CCGGCGCTGTTGCTGCTGCTGCTCAACAGACTCAATTCAGAGCAACCTGCAATGCTCAAT 1599
DB 1663 CAACAAATGTTCCGCAATTTTAAACGACCTTAAATTAAGATGCGCTGCGACCTCAAT 1722
QY 1600 CGCATATGATGGGTAGGCGTCCCTCGATACGCGCGCTGCAATGATTAAGGCGAGGTCAT 1659
DB 1723 ACAGTCTCAATGCTTTATTTAATGACCGAGCCACTCACTTATGATCCGCGAATGTACC 1782
QY 1660 GACAAAGGCTCGATCAACCAAGCGGCTGTTTTCGAATGCGGCTGCGCGAAGTTGATGCG 1719
DB 1783 GATTAAGGCAACCTCATTAAGCAATTAATCAAGAGTGTGCGCTTAAATGATGAG 1842
QY 1720 CTGTATC 1726
DB 1843 CTTTATC 1849

```

RESULT 2

AAA9722 standard; DNA; 1203 BP.

AAA9722;

21-SEP-2000 (first entry)

Pseudomonas sp. HR199 fcs-delta DNA.

Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;

vanillinic acid; flavouring; ferulic acid-coh-synthetase; fcs; fcs-delta;

db.

Pseudomonas sp.

DE19850242-A1.

04-MAY-2000.

PF 31-OCT-1998; 98DE-01050242.
XX
PR 31-OCT-1998; 98DE-01050242.
XX
PA (HAAR) HAARMANN & REIMER GMBH.
XX
PI Rabenhorst J, Steinhuechel A, Priefert H, Overhage J;
XX
DR WPI; 2000-340642/30.
XX P-PsDB; AA187972.
PT Transformed organisms, used for production of vanillin and other
XX methoxyphenols, have altered catabolism of eugenol or ferulic acid.
PS Claim 8; Fig 2i; 80pp; German.
XX
XX This invention describes novel transformed and/or mutated uni- or multi-
CC cellular organisms (A) in which enzymes (I) involved in catabolism of
CC eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that
CC the organism accumulates the intermediates conferyl alcohol (CA1),
CC conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
CC INDEPENDENT CLAIMS are also included for the following: (a) gene
CC structures in which sequences that encode one or more of the enzymes CA1-
CC dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-
CC aldolase, beta-ketochiolase, V-dehydrogenase and VA-demethylase have been
CC altered and/or inactivated; (b) vector containing at least one gene
CC structure of (a); (c) biotechnical production of organic compounds
CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
CC production method of (A). (A) are used to prepare organic compounds,
CC preferably alcohols, aldehydes and organic acids, especially CA1, CA2,
CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
CC The inexpensive starting material eugenol can be converted to V in a
CC single step. This sequence encodes the Pseudomonas sp. HR199 ferulic acid
CC -CoA-synthetase, fcs-delta which is described in the method of the
CC invention
XX
XX Sequence 1203 BP; 217 A; 337 C; 357 G; 292 T; 0 U; 0 Other;
SQ
Query Match 22.9%; Score 404.6; DB 3; Length 1203;
Best Local Similarity 91.6%; Pred. No. 3.4e-104;
Matches 428; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 ATGCGTTCTCTCGAGGCGCTTCTCCCTCCCGGATCGAATTCTTGAGCGTTCGAGCAT 60
DB 342 ATGCGTTCTCTCGAGGCGCTTCTCCCTCCCGGATCGAATTCTTGAGCGTTCGAGCAT 401
QY 61 TGGGCTTAAGACCCGTCAGAAACAACCTGCTGCTCCAGGCGGCAATGGGGAATGG 120
DB 402 TGGGCTTAAGACCCGTCAGAAACAACCTGCTGCTCCAGGCGGCAATGGGGAATGG 461
QY 121 CGTCGTATCAGCTACCGGAAATGTTCCACAACGTCGGCGCATCGACAGAGCTTGCTT 180
DB 462 CGTCGTATCAGCTACCGGAAATGTTCCACAACGTCGGCGCATCGACAGAGCTTGCTT 521
QY 181 CTTTACGCAATATCGAGAGCGCGCTGCTTATCGTCTTGGAAATGACCTGGAAAT 240
DB 522 CTTTACGCAATATCGAGAGCGCGCTGCTTATCGTCTTGGAAATGACCTGGAAAT 581
QY 241 CTTTACGCAATATCGAGAGCGCGCTGCTTATCGGCAATTCCTATTCGCCGATGCTCTGCT 300
DB 582 CTTTACGCAATATCGAGAGCGCGCTGCTTATCGGCAATTCCTATTCGCCGATGCTCTGCT 641
QY 301 TATTCATGCTGTGCGCAAGATTGGCGAAGCTGCTCATTCGTAAGTCTTTCGCAACCG 360
DB 642 TATTCATGCTGTGCGCAAGATTGGCGAAGCTGCTCATTCGTAAGTCTTTCGCAACCG 701
QY 361 GAGCTGCTCTTTCGCGCATGACAGACCTTTCAGGCGGCAATTTGAGACCATTCCTCCG 420
DB 702 GAGCTGCTCTTTCGCGCATGACAGACCTTTCAGGCGGCAATTTGAGACCATTCCTCCG 761
QY 421 GAGCTGCTCTTTCGCGCATTCTTCACTCGAGGCGAATTTGCGGCGCGCG 467
DB 762 GCGAAGTGTGATGCGGTATCGGTGGAAGATCATTCATGCTGCG 808

RESULT 3
AAA39720
ID AAA39720 standard; DNA; 2188 BP.
XX
AC AAA39720;
XX
DT 21-SEP-2000 (first entry)
XX
XX Pseudomonas sp. HR199 fcs-omega-Km DNA.
DE
XX Eugenol; ferulic acid; conferyl alcohol; conferyl aldehyde; vanillin;
KM vanillic acid; flavouring; ferulic acid-CoA-synthetase; fcs;
KM fcs-omega-Km; de.
XX
OS Pseudomonas sp.
XX
XX DE19850242-A1.
XX
PD 04-MAY-2000.
XX
XX 31-OCT-1998; 98DE-01050242.
XX
XX 31-OCT-1998; 98DE-01050242.
PR 31-OCT-1998; 98DE-01050242.
XX
PA (HAAR) HAARMANN & REIMER GMBH.
XX
PI Rabenhorst J, Steinhuechel A, Priefert H, Overhage J;
XX
DR WPI; 2000-340642/30.
XX P-PsDB; AA187970.
XX
XX Transformed organisms, used for production of vanillin and other
PT methoxyphenols, have altered catabolism of eugenol or ferulic acid.
PS Claim 8; Fig 2g; 80pp; German.
XX
XX This invention describes novel transformed and/or mutated uni- or multi-
CC cellular organisms (A) in which enzymes (I) involved in catabolism of
CC eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that
CC the organism accumulates the intermediates conferyl alcohol (CA1),
CC conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
CC INDEPENDENT CLAIMS are also included for the following: (a) gene
CC structures in which sequences that encode one or more of the enzymes CA1-
CC dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-
CC aldolase, beta-ketochiolase, V-dehydrogenase and VA-demethylase have been
CC altered and/or inactivated; (b) vector containing at least one gene
CC structure of (a); (c) biotechnical production of organic compounds
CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
CC production method of (A). (A) are used to prepare organic compounds,
CC preferably alcohols, aldehydes and organic acids, especially CA1, CA2,
CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
CC The inexpensive starting material eugenol can be converted to V in a
CC single step. This sequence encodes the Pseudomonas sp. HR199 ferulic acid
CC -CoA-synthetase, fcs-omega-Km which is described in the method of the
CC invention
XX
XX Sequence 2188 BP; 408 A; 607 C; 670 G; 503 T; 0 U; 0 Other;
SQ
Query Match 22.4%; Score 397.2; DB 3; Length 2188;
Best Local Similarity 99.3%; Pred. No. 5.6e-102;
Matches 399; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGCGTTCTCTCGAGGCGCTTCTCCCTCCCGGATCGAATTCTTGAGCGTTCGAGCAT 60
DB 342 ATGCGTTCTCTCGAGGCGCTTCTCCCTCCCGGATCGAATTCTTGAGCGTTCGAGCAT 401
QY 61 TGGGCTTAAGACCCGTCAGAAACAACCTGCTGCTCCAGGCGGCAATTTGGGGAATGG 120
DB 402 TGGGCTTAAGACCCGTCAGAAACAACCTGCTGCTCCAGGCGGCAATTTGGGGAATGG 461
QY 121 CGTCGTATCAGCTACCGGAAATGTTCCACAACGTCGGCGCATTCGACAGAGCTTGCTT 180

```

Db      462 CGTCGTATCAGCTACGCGAAATGTTCCAAACGTCGCGCCATCGACAGACTTGCTT
QY      181 CCTTACGGAATACGCGAGAGCGTCGCTATGCTCTCTGGAATGACCTGGAACAT
Db      522 CTTACGGAATACGCGAGAGCGTCGCTATGCTCTCTGGAATGACCTGGAACAT
QY      241 CTTACGGAATACGCGAGAGCGTCGCTATGCTCTCTGGAATGACCTGGAACAT
Db      582 CTTACGGAATACGCGAGAGCGTCGCTATGCTCTCTGGAATGACCTGGAACAT
QY      301 TATTCACTGCTGCGCAAGATTGGGGAAGCTGCTGACATGCTAGTCTTTGCAACCG
Db      642 TATTCACTGCTGCGCAAGATTGGGGAAGCTGCTGACATGCTAGTCTTTGCAACCG
QY      361 GGACTGCTCTTGTGCTGCGCAAGATTGGGGAAGCTGCTGACATGCTAGTCTTTGCAACCG
Db      702 GGACTGCTCTTGTGCTGCGCAAGATTGGGGAAGCTGCTGACATGCTAGTCTTTGCAACCG

```

RESULT 4

AAA9721
ID AAA9721 standard; DNA; 2171 BP.

AC AAA9721;

DT 21-SEP-2000 (first entry)

DE Pseudomonas sp. HR199 fcs-omega-Gm DNA.

XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
KW vanillinic acid; flavouring; ferulic acid-CoA-synthetase; fcs;
KM fcs-omega-Gm; ds.

XX Pseudomonas sp.

OS DE19850242-A1.

PN 04-MAY-2000.

PD 31-OCT-1998; 98DE-01050242.

PE 31-OCT-1998; 98DE-01050242.

PR (HAAR) HAARMANN & REIMER GMBH.

PA Rabenhorst J, Steinhuechel A, Priefert H, Overhage J;

PI WPI; 2000-340642/30.

DR P-PSDB; AAY87971.

XX Transformed organisms, used for production of vanillin and other
XX methoxyphenols, have altered catabolism of eugenol or ferulic acid.

PS Claim 8; Fig 2h; 80pp; German.

XX This invention describes novel transformed and/or mutated uni- or multi-
CC cellular organisms (A) in which enzymes (I) involved in catabolism of
CC eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that
CC the organism accumulates the intermediates coniferyl alcohol (CAI),
CC coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillinic acid (VA).
CC INDEPENDENT CLAIMS are also included for the following: (a) gene
CC structures in which sequences that encode one or more of the enzymes CAI-
CC dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-
CC aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase have been
CC altered and/or inactivated; (b) vector containing at least one gene
CC structure of (a); (c) biotechnical production of organic compounds
CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
CC production method of (A). (A) are used to prepare organic compounds,
CC preferably alcohols, aldehydes and organic acids, especially CAI, CA2,
CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
CC The inexpensive starting material eugenol can be converted to V in a
CC single step. This sequence encodes the Pseudomonas sp. HR199 ferulic acid

CC -CoA-synthetase, fcs-omega-Gm which is described in the method of the
CC invention
XX Sequence 2171 BP; 443 A; 591 C; 616 G; 521 T; 0 U; 0 Other;

Query Match 22.4%; Score 396.8; DB 3; Length 2171;
Best Local Similarity 99.5%; Pred. No. 7.2e-102;
Matches 398; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY      1 ATGCGTCTCTGAGGCGCTTCTCCCTCCCGGGTGCAGATTCTTGAGCGTCCAGCAT
Db      342 ATGCGTCTCTGAGGCGCTTCTCCCTCCCGGGTGCAGATTCTTGAGCGTCCAGCAT
QY      61 TGGGCTAAGACCCGTCAGAAACAAACCTGCTGCTGCGAGGCGGCAATGGGAATGG
Db      402 TGGGCTAAGACCCGTCAGAAACAAACCTGCTGCTGCGAGGCGGCAATGGGAATGG
QY      121 CGTCGTATCAGCTACCGGGAATGTTCCCAACAGTCCGGCCATGCGACAGACTTGCTT
Db      462 CGTCGTATCAGCTACCGGGAATGTTCCCAACAGTCCGGCCATGCGACAGACTTGCTT
QY      181 CCTTACGGAATACGCGAGAGCGTCGCTATGCTCTGGAATGACCTGGAACAT
Db      522 CCTTACGGAATACGCGAGAGCGTCGCTATGCTCTGGAATGACCTGGAACAT
QY      241 CTTACGGAATACGCGAGAGCGTCGCTATGCTCTGGAATGACCTGGAACAT
Db      582 CTTACGGAATACGCGAGAGCGTCGCTATGCTCTGGAATGACCTGGAACAT
QY      301 TATTCACTGCTGCGCAAGATTGGGGAAGCTGCTGACATGCTAGTCTTTGCAACCG
Db      642 TATTCACTGCTGCGCAAGATTGGGGAAGCTGCTGACATGCTAGTCTTTGCAACCG
QY      361 GGACTGCTCTTGTGCTGCGCAAGATTGGGGAAGCTGCTGACATGCTAGTCTTTGCAACCG
Db      702 GGACTGCTCTTGTGCTGCGCAAGATTGGGGAAGCTGCTGACATGCTAGTCTTTGCAACCG

```

RESULT 5

AAA9731
ID AAA9731 standard; DNA; 1543 BP.

AC AAA9731;

DT 21-SEP-2000 (first entry)

DE Pseudomonas sp. HR199 vdh-delta DNA.

XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
KW vanillinic acid; flavouring; vanillin dehydrogenase; vdh; vdh-delta; ds.
KM

XX Pseudomonas sp.

OS DE19850242-A1.

PN 04-MAY-2000.

PD 31-OCT-1998; 98DE-01050242.

PE 31-OCT-1998; 98DE-01050242.

PR (HAAR) HAARMANN & REIMER GMBH.

PA Rabenhorst J, Steinhuechel A, Priefert H, Overhage J;

PI WPI; 2000-340642/30.

DR P-PSDB; AAY87981.

XX Transformed organisms, used for production of vanillin and other
XX methoxyphenols, have altered catabolism of eugenol or ferulic acid.

PS Claim 8; Fig 2r; 80pp; German.

CC This invention describes novel transformed and/or mutated uni- or multi-cellular organisms (A) in which enzymes (I) involved in catabolism of eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that the organism accumulates the intermediates conferyl alcohol (CA1), conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA). CC INDEPENDENT CLAIMS are also included for the following: (a) gene structures in which sequences that encode one or more of the enzymes CA1-dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-aldolase, beta-ketochiolase, V-dehydrogenase and VA-demethylase have been altered and/or inactivated; (b) vector containing at least one gene structure of (a); (c) biotechnical production of organic compounds, particularly alcohols, aldehydes or organic acids using (A); and (d) production method of (A). (A) are used to prepare organic compounds, preferably alcohols, aldehydes and organic acids, especially CA1, CA2, FA, V and/or VA, most particularly V, a well-known natural flavouring. CC The inexpensive starting material eugenol can be converted to V in a single step. This sequence encodes the Pseudomonas sp. HR199 vanillin dehydrogenase, vdh-delta which is described in the method of the invention

SO Sequence 1543 BP; 306 A; 387 C; 473 G; 377 T; 0 U; 0 Other;

Query Match 4.1%; Score 73; DB 3; Length 1543;

Best Local Similarity 100.0%; Pred. No. 6.7e-10;

Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1698 GCGGTCGCGCAAGTTGATCGCTGTATCGTGTGAAGATCAATCCATGCTGCTGACGA 1757

DB 8 GCGGTCGCGCAAGTTGATCGCTGTATCGTGTGAAGATCAATCCATGCTGCTGACGA 67

OY 1758 GGCACACACTGTGA 1770

DB 68 GGCACACACTGTGA 80

RESULT 6

AAA39730 standard; DNA; 2509 BP.

AC AAA39730;

DT 21-SEP-2000 (first entry)

DE Pseudomonas sp. HR199 vdh-omega-Gm DNA.

KW Eugenol; ferulic acid; conferyl alcohol; conferyl aldehyde; vanillin;

KM vanillic acid; flavouring; vanillin dehydrogenase; vdh; vdh-omega-Gm; de.

OS Pseudomonas sp.

PN DE19850242-A1.

PD 04-MAY-2000.

PF 31-OCT-1998; 98DE-01050242.

PR 31-OCT-1998; 98DE-01050242.

PA (HAAR) HAARMANN & REIMER GMBH.

PI Rebenhorst J, Steinhuechel A, Priefert H, Overhage J;

DR WPI; 2000-340642/30.

DR P-PSDB; AAY87980.

PT Transformed organisms, used for production of vanillin and other methoxyphenols, have altered catabolism of eugenol or ferulic acid.

PS Claim 8; Fig 2q; 80pp; German.

CC This invention describes novel transformed and/or mutated uni- or multi-cellular organisms (A) in which enzymes (I) involved in catabolism of

CC eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that the organism accumulates the intermediates conferyl alcohol (CA1), conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA). CC INDEPENDENT CLAIMS are also included for the following: (a) gene structures in which sequences that encode one or more of the enzymes CA1-dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-aldolase, beta-ketochiolase, V-dehydrogenase and VA-demethylase have been altered and/or inactivated; (b) vector containing at least one gene structure of (a); (c) biotechnical production of organic compounds, particularly alcohols, aldehydes or organic acids using (A); and (d) production method of (A). (A) are used to prepare organic compounds, preferably alcohols, aldehydes and organic acids, especially CA1, CA2, FA, V and/or VA, most particularly V, a well-known natural flavouring. CC The inexpensive starting material eugenol can be converted to V in a single step. This sequence encodes the Pseudomonas sp. HR199 vanillin dehydrogenase, vdh-omega-Gm which is described in the method of the invention

SO Sequence 2509 BP; 532 A; 640 C; 731 G; 606 T; 0 U; 0 Other;

Query Match 4.1%; Score 73; DB 3; Length 2509;

Best Local Similarity 100.0%; Pred. No. 8.2e-10;

Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1698 GCGGTCGCGCAAGTTGATCGCTGTATCGTGTGAAGATCAATCCATGCTGCTGACGA 1757

DB 8 GCGGTCGCGCAAGTTGATCGCTGTATCGTGTGAAGATCAATCCATGCTGCTGACGA 67

OY 1758 GGCACACACTGTGA 1770

DB 68 GGCACACACTGTGA 80

RESULT 7

AAA39729 standard; DNA; 2526 BP.

AC AAA39729;

DT 21-SEP-2000 (first entry)

DE Pseudomonas sp. HR199 vdh-omega-Km DNA.

KW Eugenol; ferulic acid; conferyl alcohol; conferyl aldehyde; vanillin;

KM vanillic acid; flavouring; vanillin dehydrogenase; vdh; vdh-omega-Km; de.

OS Pseudomonas sp.

PN DE19850242-A1.

PD 04-MAY-2000.

PF 31-OCT-1998; 98DE-01050242.

PR 31-OCT-1998; 98DE-01050242.

PA (HAAR) HAARMANN & REIMER GMBH.

PI Rebenhorst J, Steinhuechel A, Priefert H, Overhage J;

DR WPI; 2000-340642/30.

DR P-PSDB; AAY87979.

PT Transformed organisms, used for production of vanillin and other methoxyphenols, have altered catabolism of eugenol or ferulic acid.

PS Claim 8; Fig 2p; 80pp; German.

CC This invention describes novel transformed and/or mutated uni- or multi-cellular organisms (A) in which enzymes (I) involved in catabolism of eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that the organism accumulates the intermediates conferyl alcohol (CA1),

CC conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
CC INDEPENDENT CLAIMS are also included for the following: (a) gene
CC structures in which sequences that encode one or more of the enzymes CA1-
CC dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-
CC aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase have been
CC altered and/or inactivated; (b) vector containing at least one gene
CC structure of (a); (c) biotechnical production of organic compounds
CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
CC production method of (A). (A) are used to prepare organic compounds,
CC preferably alcohols, aldehydes and organic acids, especially CA1, CA2,
CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
CC The inexpensive starting material eugenol can be converted to V in a
CC single step. This sequence encodes the pseudomonas sp. HR199 vanillin
CC dehydrogenase, vdh-omega-Km which is described in the method of the
CC invention
XX
SQ Sequence 2526 BP; 497 A; 656 C; 785 G; 588 T; 0 U; 0 Other;
Query Match 4.1%; Score 73; DB 3; Length 2526;
Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1698 GCGGTGGCGCAAGTGCATGCTGATCGTGGAAGATCATCCATGCTGGGTACGA 1757
DB 8 GCGGTGGCGCAAGTGCATGCTGATCGTGGAAGATCATCCATGCTGGGTACGA 67
QY 1758 GGGCAGACTGTGA 1770
DB 68 GGGCAGACTGTGA 80
Db 68 GGGCAGACTGTGA 80
RESULT 8
AA064205
ID AA064205 standard; cDNA; 1879 BP.
XX
AC AA064205;
XX
DT 25-MAR-2003 (revised)
DT 18-NOV-1994 (first entry)
XX
DE snbA gene encoding enzyme in streptogramin biosynthetic pathway.
XX
KW Antibiotic; streptogramin; snbA; snab; snac; biosynthesis; enzyme;
KW biosynthetic pathway; Streptomyces pristinaespiralis; snbA; snbR; ds.
XX
OS Streptomyces pristinaespiralis.
XX
FH Key Location/Qualifiers
FT CDS 110..1858
FT CDS /*tag= a
XX
PN FR2696189-A1.
XX
PD 01-APR-1994.
XX
PF 25-SEP-1992; 92FR-00011441.
XX
PR 25-SEP-1992; 92FR-00011441.
XX
PA (RHON) RHONE POULENC RORER SA.
XX
PI Blanc V, Blanche F, Crouzet J, Jacques N, Lacroix P, Thibaut D;
PI Zagorec M;
XX
DR WPI: 1994-128286/16.
DR P-PSDB; AARS4205.
XX
PT DNA involved in streptogramin antibiotic biosynthesis - for prodn. or bio
PT -conversion of streptogramin(s) or prodn. of streptogramin intermediates,
PT derivs. or hybrid antibiotics.
XX
PS Claim 2; Page 57-60; 83pp; French.
XX

CC The snbA gene product is involved in the biosynthesis of streptogramins,
CC antibiotics active against Gram-positive bacteria. The identification of
CC the sequences encoding the enzymes involved in the biosynthetic pathway
CC means that they can be isolated and manipulated. Mutant microorganisms in
CC which a step in the streptogramin biosynthetic pathway is blocked can be
CC cultured to produce streptogramin intermediates, which may later be
CC converted to streptogramin derivatives. Recombinant cells may also be
CC used for the bioconversion of streptogramins from one form to another or
CC for the production of hybrid antibiotics. (Updated on 25-MAR-2003 to
CC correct PN field.)
XX
SQ Sequence 1879 BP; 235 A; 682 C; 720 G; 242 T; 0 U; 0 Other;
Query Match 3.2%; Score 57.4; DB 2; Length 1879;
Best Local Similarity 56.7%; Pred. No. 2e-05;
Matches 106; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 1071 CACCGAACCCTGTCATGCGCTGCTTACATTTGGGCTCCAGACCTGGCTGCAGATCAA 1130
DB 1186 CACCGAAGGCGCGCCGCTGTCGAGGCCGACGATACGCTGCGACCCGACGCGCG 1245
QY 1131 GCTCGTTCCGCTGATGAGGAAATTGAGAGGCGCTTCCATGTCGCGACGTCATGAGCGG 1190
DB 1246 GCGCGTGGCCCGCGGTGAAGCCGCTAACCTCACCCTCGGCTTACAGCTGGCGG 1305
QY 1191 CTACTGCGCTGCTCTTGAACAAATGCCCAGAGCTTCAGAGAGAAAGCTATTACTGCTC 1250
DB 1306 CTACTACCGGGCCCCCGAGACAAACGCCGCGCTTACCGAGACGCTTACCGCAG 1365
QY 1251 CGGTGAT 1257
DB 1366 CGCGGAT 1372
Db 1366 CGCGGAT 1372
RESULT 9
ADA71938
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-1B001105.
XX
PR 22-JUN-2001; 2001WO-1B001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Qian S, Tao Y, Whitnam S, Xie Z, Zhu T, Zou G;
XX
DR WPI: 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1

comprises identifying a gene whose expression is significantly altered in the incompatible interaction of plant gene expression relative to the expression of the gene in an uninfected plant, in a mutant plant that does not express a gene associated with response to pathogenic infection, or in a corresponding incompatible or compatible interaction. (M) is useful for conferring resistance to resistance or tolerance to a plant to bacterial, fungal or viral infection. The present sequence was used to illustrate the invention.

SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 3.0%; Score 53.2; DB 8; Length 2000;
 Best Local Similarity 8.9%; Pred. No. 0.00032;
 Matches 66; Conservative 337; Mismatches 335; Indels 4; Gaps 1;

Oy	886	GAGCGAGACAGTACCTCCGCGAAGCGCTTCTTCGCTCCAGTAAGACTGTTCTTTCGCG	945
Db	18	RRRMTAAMMSICARNGSSRMSKMSMSISRYRCCSSCGAKIMTTRKSKMYSSASAGTG	77
Oy	946	GCGCGTGGGTTGTCCCAAGAGGATCTGGATCGTTTGAACCGGGTGCCTGAACAGACTGT	1008
Db	78	SKMSGSGTSGKMKKRYRKRKRGRKGRMRSMRMGRRRCARSGRMAGSGRMVG	137
Oy	1006	GGTGAGCGCATTCGATGATGCGGGGTCTGGGCATGACGGAAGACTGCTTTCGCACT	1065
Db	138	GKSRRMSYMMCYARCGSGCKRKKSKGSGWGKTCSRAGARGSGMSGAKYKSGMSKRMVM	197
Oy	1066	TTTACACCGGACCGGCTGCATGCTGTGTTACATTGGGCTCCAGCGCTGCTGCGCAG	1122
Db	198	SSCGRSGGGRSAYSVRYGTSSKTYTKMYTYSASRCMRAYMTTYSISWAGCSYMWCSK	257
Oy	1126	GTCAGACTCGTTCGCGTGATGGAATAATTGAAGAGCGCTTTCATGATCGGACGCTCATG	1188
Db	258	RRSMMKMKMRKRMSRSGTSMSTKMMCTPAKYSYRKCCTMYTRGGMRBATYMGRG	317
Oy	1186	AGCGGCTTAGCGCGTCTCTGGAACAAATGCCCCAAGCGTTGACGAGGAAGCTATTAC	1245
Db	318	YMSRMAAMYKKMYRYRGYKMGKRMGWMAGRMHRSMCRMSKACYRYMRMRMTBRBRMAK	377
Oy	1246	TGCTCCGGTGATGCATCAATTTGGCAAGTCTCGCCGATCTCAGAAAAGTGTGATGTTT	1300
Db	378	KSSRTSRKKRKMCRKRRKRYKMRYSRRSCKRARMKRCBSGAMKMGKCMCTCRWK	437
Oy	1306	GACGGTCAATTCTGAAGACTTCAAGCTGTCTCAGGGGTATTGTACGCGTTGGGCA	1365
Db	438	SYGMMRMKMSKRMASKTKMMSMTMYRRKKKCSRTTMGKTRGMMGTMRCACTYKRSQMK	497
Oy	1366	TTGCGCACGCGGCGGCTTCTGGAAGCGGCTTTCAGTCTGGAAGTGAAGTGTCTGCT	1421
Db	498	RKCRRRRRGRMYRBMWKRYMSABRTMYRCAKXYSASAARACRMYGKGYVMAGMMVK	557
Oy	1426	CCTGATGCTGAATGCGCTTGATGCTCGTGTTCGCGCTTTCGACTGCTGCGGCTTTG	1488
Db	558	RYKRYMYMKMMMYRKYKSKSWYKMSYVASCMKSARCAKMKCKRSMMSKSM--	614
Oy	1486	TCGGGGCTAGAGAAAAGAGGCGTCGACGCGCGAGGTGCTTGCCAGTGAAGCGCGATCGGGC	1545
Db	615	-RSSRKKCKAKSKSSAKRYAMMGGTTSGSRMSRMSSTCYRKMGSMKSTTMYMYSK	673
Oy	1546	TGGTTTGTCTACTGCTCAAAAGCATCTATCGAAGAACAACTGGCAATGCCAGTGCATC	1605
Db	674	YTYAAYGSGYMYRBYAMCMYMMRWYYRYRBSYMTYMAWTSTBTAMTMGMKXSGYMYTSM	733
Oy	1606	ATGTGGGTAGGGCTCTCTCGATA	1627
Db	734	YKYCKCSWKYSRSMYYWSMMWA	755

```
RESULT 10
ADB68842
ID      ADB68842 standard; DNA; 536 BP
XX
AC      ADB68842;
```

XX 04-DEC-2003 (first entry)
DT
XX
XX Minority luxI consensus sequence DNA 14.
DE
XX
XX quorum sensing; lux homologue; luxI; ds.
KW
XX
XX Unidentified.
OS
XX
XX WO2003057902-A2.
PN
XX
XX 17-JUL-2003.
PD
XX
XX 08-JAN-2003; 2003WO-US000479.
PF
XX
XX 08-JAN-2002; 2002US-0346531P.
PR
XX
XX 07-JAN-2003; 2003US-00338110.
PP
XX
XX (FRAU) FRAUNHOFER USA INC.
PA
XX
XX Fuhrmann JJ, Romesser JA;
PI
XX
XX WPI; 2003-618102/58.
DR
XX
XX
PT Detecting quorum sensing potential of a Gram-negative bacterium in a
PT sample comprises performing a polymerase chain reaction using nucleic
PT acids extracted from a sample containing a microorganism.
PS
XX
XX Disclosure; Fig 10; 86pp; English.
XX
XX The invention relates to a novel method for detecting the quorum sensing
CC potential of a microorganism in a sample which comprises performing PCR
CC using nucleic acids extracted from a sample containing at least one type
CC of microorganism. The method may be useful for detecting the quorum
CC sensing potential of a microorganism in a sample by amplifying a fragment
CC of a lux gene or homologue. The current sequence is that of the minority
CC luxI consensus sequence DNA of the invention.
XX
SQ Sequence 536 BP; 20 A; 52 C; 50 G; 40 T; 0 U; 374 Other;
XX

	Query Match	2.6%	Score 46.8;	DB 10;	Length 536;		
	Best Local Similarity	11.7%;	Pred. No. 0.012;				
	Matches	54;	Conservative	220;	Mismatches	187;	
				Indels	0;	Gaps	0
Qy	382	GCAGCACCTTTCACGCGCGCAATTGAGCCATTTCTGCCGACGACGTGCCGGCATCTTC	441				
	:	:::::	:	:	:	:	
	:	:	:	:	:	:	
	:	:	:	:	:	:	
Db	73	SVRSBYCVWYSKNNVVGVDMMSBKVMGARKBGAYCARTCGAYMYVSMSASHRCSKC	132				
Qy	442	ACTGAGCGCGAAATTGGCCCCGGCGCGACGGTAGTTTGAACGCTGCTGGACAGCT	501				
	:	:	:	:	:	:	
	:	:	:	:	:	:	
	:	:	:	:	:	:	
Db	133	TACTVSBTSGSCHKRWGNVSYYVBDAHTSKRYGHTHYGSCSCTGHTDCBACSACV	192				
Qy	502	GATGGGATTTGAGCGAGATAATGCTTTTGCGGCAACTGGCCCCGATACGATTTGCCAAGTTC	561				
	:	:	:	:	:	:	
	:	:	:	:	:	:	
	:	:	:	:	:	:	
Db	193	BDSGBMAVCTGCTBRAKAVNSTYTTTHSSBMYNYSBKCRBVSRYRVMMYRCISNVSAT	252				
Qy	562	TTGTTCACCTTCTGCGTCTAACCAATGCGCTTAAGCGGATGCCGACTACTACAGCAATGCTC	621				
	:	:	:	:	:	:	
	:	:	:	:	:	:	
	:	:	:	:	:	:	
Db	253	CCVKMSRINTGGGAHRTMCNCGMTWCSCSRCHRSASAMYGKCSRTGMKMTTY	312				
Qy	622	TGCGGCATTCAGACGATGCTTCTGCAACTTTCGCGGTTTTGGTGAAGACCCGCGGTG	681				
	:	:	:	:	:	:	
	:	:	:	:	:	:	
	:	:	:	:	:	:	
Db	313	TKSDVDVRBCSTSVARMSMGCSBSYVWSBGNGCYVRREVSNTSTSSVRSRSM	372				
Qy	682	CTGTGAGCTGTGCTGGCCGTGAACAACCACTTTGCGCGGACGACACACATCGGACTGCTG	741				
	:	:	:	:	:	:	
	:	:	:	:	:	:	
	:	:	:	:	:	:	
Db	373	HCRBRSSRATGAWWGSGYMKWTMCDVNKMSAACGBRITMDBBHSANGVAVYGGBYGS	432				
Qy	742	TTGTACACAGCGCGGACGTACTACCTTACACACGGTAAACCAACGCCCAAGGTTCCGC	801				
	:	:	:	:	:	:	
	:	:	:	:	:	:	
	:	:	:	:	:	:	
Db	433	YBNHSVMMHVBRHRVDRMHRYTSTSGCSITSYGGAYCRVADCRMDSCYVVRNSVBY	492				
Qy	802	GAGACGCTTGCAACTTGAGCGCAATCTCTCCACTGCGTA	842				

Db 493 KYWSCKYYRSSMTHYCVNRHBBGVBNMVKYKBVGVGVDREBYKV 533

RESULT 11
AAZ23757
ID AAZ23757 standard; DNA; 5057 BP.

DT	27-AUG-2003	(revised)
DT	14-JAN-2000	(first entry)

P. fluorescens DNA encoding ICS orfa and orfd.

KM isochlorismate synthase; ICS; plant; pathogen resistance; chitinase;
KM pathogen inducible promoter; antipathogenic protein; toxin;
KM antifungal protein, albumin-type protein; hypersensitive response;
KM isochlorismate pyruvate synthase; ss.

OS *Pseudomonas fluorescens*.

FH	Key	Location/Qualifiers
FT	CDS	
		207. .1382

FT	/product="isochorismate synthase orfA"
FT	4516.4851
FT	/*tag= b
FT	/product="isochorismate pyruvate lyase orfD"

PN WO9950423-A2.

PD 07-OCT-1999

PF 25-MAR-1999; 99WO-EP002176.

PR	31-MAR-1998;	98US-0080203P.
PR	03-APR-1998;	98US-0080625P.

PA	(MOGE-)	MOGEN INT NV.
PA	(UYLE-)	RIJKSUNIV LEIDEN.
PA	(UYNI-)	UNIV NIJMEGEN.

PI Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH;
PI Van Tegelen LJP, Willems GJ, Croes AF, Stuijver MH, Custers J,
PI Simons LH, Melchers LS, Bol JF;

DR WPI; 1999-610856/52.

P-PSDB; AAY50336, AAY50337, AAY50338, AAY50339.

PT Method for inducing pathogen resistance in plants.

PS Claim 5; Page 54-58; 66pp; English.

CC This invention describes a novel method for the induction of pathogen
CC resistance in plants, by transformation with an expression cassette
CC harboring a gene coding for an isochlorismate synthase (ICS). A pathogen
CC inducible promoter can be used to drive expression of a heterologous
CC protein. The heterologous protein used in the method of the invention is
CC an anti-pathogenic protein e.g. chitinase, glucanase, osmotin, lectin,
CC ascarichate oxidase, oxalate oxidase, metalains, toxins from *Bacillus*
CC *thuringiensis*, or antifungal proteins isolated from *Mirabilis jalapa*,
CC *Amaranthus*, *Raphanus*, *Brassica*, *Sinapis*, *Arabidopsis*, *Dahlia*, *Chnicus*,
CC *Lathyrus*, *Clitoria*, *Allium* seeds, *Aralia* and *Impatiens* and albumin-type
CC proteins, such as thionine, napin, barley trypsin inhibitor, cereal
CC gliadin and wheat- α -amylase, or a protein that can induce a
CC hypersensitive response, such as Cf, Bs3 and Pc proteins from tomato and
CC N-protein from tobacco. This sequence encodes the Pseudomonas fluorescens
CC isochlorismate synthase *orfA* protein and the isochlorismate pyruvate lyase
CC *orfD* protein which are described in the method of the invention. (updated
CC on 27-AUG-2003 to correct OS field.)

Sequence 5057 BP; 1171 A; 1554 C; 1302 G; 1030 T; 0 U; 0 Other;

Query March	2.6%	Score	46.6	DB	2	Length	5057
Best Local Similarity	66.3%	Pred. No.	0.035				
Matches 67; Conservative	0	Mismatches	34	Indels	0	Gaps	0

Qy	Db
1168	CGATGGTCCGACGTCATAGCGGCTACGCGCGCTCCTGAA
1222	CAAAATGCCAAGCGTTC
2570	CGGGCCCTTACACTTTTGGCGGTACTACCAAGCCCGAACA
2628	AAATGCCACGGTTC

Qy	1228	GACGAGGAAGGCTATTACTGCTCCGGTGATGCCATCAAATT	1268
Db	2630	GATTAAGAGGGTACTACTCCGGCGACCTCGTCCAAC	2670

RESULT 12
ACA42560
ID ACA42560 standard; DNA; 1644 BP

AC ACA42560 ;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #24217.

KW Antisense; ds; prokaryotic essential gene; cell proliferation,
KW drug design; gene.

***Pseudomonas aeruginosa*.**

PN WO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242
PR 06-SEP-2001; 2001US-00948993
PR 25-OCT-2001; 2001US-0342923P
PR 08-FEB-2002; 2002US-00072851
PR 06-MAR-2002; 2002US-0362699P

PA (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

DR WPI; 2003-029926/02.

DR P-PSDB; ABU38690.

PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 30430; 1766pp; English

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene

DT 02-MAY-2001 (first entry)
 XX Pseudomonas sp lipopeptide synthase DNA ORF08563.
 DE
 XX
 XX Lipopeptide synthase; transgenic plant; antiviral; antibacterial;
 KM antifungal; surfactant; microbial depollution; emulsifier;
 KW tertiary crude oil recovery; ORF08563; ds.
 XX
 XX Pseudomonas sp.
 OS
 XX
 XX WO200107589-A2.
 PN
 XX
 XX 01-FEB-2001.
 PD
 XX
 XX 21-JUL-2000; 2000MO-EP007002.
 PF
 XX
 XX 27-JUL-1999; 99DE-01035106.
 PR
 XX
 XX (TIGR-) TIGR INST GENOMIC RES.
 PA (QIAG-) QIAGEN GMBH.
 PA (GBFB-) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
 PA (MED1-) MEDIZINISCHE HOCHSCHULE HANNOVER.
 XX
 XX Frazer CM, Tuemmler B, Hoheisel J, Duesterhoeft A, Hilbert H;
 PI Timms KN, Moore E, Straetz M, Heim S, Golyshin P;
 XX
 XX WPI; 2001-159715/16.
 DR
 XX
 XX New DNA encoding a bacterial lipopeptide synthase, useful e.g. for
 PT producing antimicrobial agents, surfactants with increased
 PT bioavailability in microbial depollution methods, and transgenic plants.
 PS
 XX
 XX Claim 3a; Page 28-30; 33pp; German.
 XX
 CC This invention describes a novel DNA sequence (I) that expresses products
 CC having the biological function of lipopeptide synthase. The invention
 CC also describes (1) recombinant expression vectors containing (I); (2)
 CC prokaryotic and eukaryotic cells transformed or transfected with (I) or
 CC the vector of (2); (3) production of lipopeptide synthases by culturing
 CC cells of (2); (4) expression products (II) of (I), and synthetic proteins
 CC or peptides with the same sequences; (5) mono- or polyclonal antibodies
 CC (Ab) specific for (II); (6) hybridoma cells that produce monoclonal Ab;
 CC and (7) transgenic plants that contain cells of (2). (I), and their
 CC fragments, are useful for expression of recombinant lipopeptide
 CC synthases, and as probes and primers for detection, isolation and
 CC amplification of full-length cDNA sequences. (II) are used to produce
 CC transgenic plants. Lipopeptide synthases are useful for production of
 CC antiviral, antibacterial or antifungal lipopeptides, particularly useful
 CC for treatment of plants, and for production of lipopeptide surfactants
 CC used for increasing bioavailability in microbial depollution processes
 CC (acting as emulsifiers) and in tertiary crude oil recovery. Individual
 CC subunits from different lipopeptide synthases may be combined to allow
 CC synthesis of many different biologically active substances
 CC
 XX
 XX Sequence 5451 BP; 937 A; 1766 C; 1807 G; 941 T; 0 U; 0 Other;
 SQ
 Query Match 2.6%; Score 45.4; DB 5; Length 5451;
 Best Local Similarity 52.4%; Pred. No. 0.08; Indels 0; Gaps 0;
 Matches 100; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
 QY 455 TGGCCGGGGCGGCACGCGTGGTGGTGGACAGCCTGTGAGCAGCCTGTGGGATTGAGG 514
 Db 2132 TGGCGGGTGGCTGGCTGGCTGACCGAGGTGGCTTCGACGGCGTGTGGAGAGCGGTGACG 2191
 QY 515 CAATATATGCTTTGGCGGCACTGGCCCGCATACGATGGCAAGTTTGTTCACCTTCTG 574
 Db 2192 CTTCGCAACCGGCAAGTGGCCATGACCCCTGCACAGCGTGGCTTACATATTAACCCCTGG 2251
 QY 575 GCTCTACCAACTGCTTAGCGGTCGCGACTACTGACGGAATGCTCTGGCCCAATGAGC 634
 Db 2252 GCTTCACCGGGCGGCCCAAGGGTGTGGCGATACCCAGGGCGCATGAGCCAGCATGCCC 2311

QY 635 AGATGCTTCTG 645
 Db 2312 AGGTGTCGCTG 2322
 RESULT 15
 ID AC43663
 AC43663 standard; DNA; 6522 BP.
 AC43663;
 AC43663;
 AC 19-JUN-2003 (first entry)
 AC 19-JUN-2003 (first entry)
 DE Prokaryotic essential gene #25320.
 DE Prokaryotic essential gene #25320.
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 KM
 XX Pseudomonas putida.
 OS
 XX
 XX WO200277183-A2.
 PN
 XX
 XX 03-OCT-2002.
 PD
 XX
 XX 21-MAR-2002; 2002MO-US009107.
 PF
 XX
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342823P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
 PI Wall D, Tawick JD, Carr GI, Yamamoto R, Forsyth RA, Xu HH;
 XX
 XX WPI; 2003-029926/02.
 DR P-SDB; ABU39793.
 DR
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids, required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 PS
 XX
 XX Claim 14; SEQ ID NO 31533; 1766bp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2005, 15:49:18 / Search time 7360 Seconds
(without alignments)
11372.659 Million cell updates/sec

Title: US-09-750-986D-29

Perfect score: 1770

Sequence: 1 atgcgtctctcgcagcgctc.....gtgcagcgccacacgttga 1770

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapect 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

GenEmbl: 1: gb ba: *
2: gb hty: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pac: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1770	100.0	1770	6 A92118	A92118 Sequence 29
2	1770	100.0	4166	6 PSP338746	AJ238746 Pseudomon
3	1770	100.0	32679	6 A92090	A92090 Sequence 1
4	987.6	55.8	9477	1 PFL536325	AJ536325 Pseudomon
5	965.2	54.5	310029	1 AE016866	AE016866 Pseudomon
6	898.2	50.7	301214	1 AE016786	AE016786 Pseudomon
7	763.2	43.1	4407	1 PPUS56324	AJ56324 Pseudomon
8	741.4	41.9	189050	1 AL646077	AL646077 Ralstonia
9	527.8	29.8	346879	1 BX572598	BX572598 Rhodospira
10	489.6	27.7	299950	1 AP005963	AP005963 Bradyrhiz
11	445.4	25.2	1902	6 AR319843	AR319843 Sequence
12	442.8	25.0	45324	1 ACCPCAOP	LO5770 Accinobact
13	441.2	24.9	110000	6 CR543861_17	AX024534 Sequence
14	404.6	22.6	10173	1 AE008067	AE008067 Agrobacte
15	399.4	22.6	10173	1 AE009102	AE009102 Agrobacte
16	397.2	22.4	2188	6 AX024541	AX024541 Sequence
17	396.8	22.4	2171	6 AX024537	AX024537 Sequence
18	375.6	21.2	299450	1 AP005938	AP005938 Bradyrhiz

C	20	342.8	19.4	299950	1	AP005963	AP005963 Bradyrhiz
	21	285.2	16.1	7171	1	AF302797	AF302797 Pseudomon
	22	232.6	13.1	10306	1	AE005712	AE005712 Caulobact
C	23	225.6	12.7	12606	1	AE005995	AE005995 Caulobact
	24	73	4.1	1543	6	AX024490	AX024490 Sequence
	25	73	4.1	2509	6	AX024493	AX024493 Sequence
	26	73	4.1	2526	6	AX024497	AX024497 Sequence
C	27	61	3.4	125020	9	AF429315	AF429315 Homo sapi
	28	58.2	3.3	125020	9	AF429315	AF429315 Homo sapi
	29	57.4	3.2	1879	1	SPSNBAGEN	X98515 S. prietinae
	30	57.4	3.2	1879	6	A37835	A37835 Sequence 5
	31	57.4	3.2	1879	6	AR069873	AR069873 Sequence
	32	57.4	3.2	1879	6	AR099269	AR099269 Sequence
	33	57.4	3.2	1879	6	AR124154	AR124154 Sequence
	34	57.4	3.2	1879	6	AR442761	AR442761 Sequence
	35	53.2	3.0	2000	6	AX655393	AX655393 Sequence
C	36	53	3.0	302007	1	SC0939132	AL939132 Streptomy
	37	52.4	3.0	6629	1	AB072568	AB072568 Streptomy
C	38	49.4	2.8	247450	1	AP005349	AP005349 Vibrio vu
	39	48.6	2.7	310174	1	AE016870	AE016870 Pseudomon
C	40	47.8	2.7	301370	1	AE016811	AE016811 Vibrio vu
	41	46.8	2.6	12730	1	AE005055	AE005055 Halobacte
	42	46.6	2.6	5057	1	PFFBSCXB	X90356 Halobacte
	43	46.4	2.6	4662	1	PAX82644	X82644 Pseudomonas
	44	46.4	2.6	10481	1	AE004840	AE004840 Pseudomon
	45	46.2	2.6	10397	1	AE011895	AE011895 Xanthomon

ALIGNMENTS

RESULT 1
LOCUS A92118
DEFINITION Sequence 29 from Patent EP0845532.
ACCESSION A92118
VERSION A92118.1 GI:6740914
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1770)
Prieffert, H.D. and Rabenhorst, J.D.
Enzymes for the synthesis of conferyl alcohol, conferyl aldehyde, ferulic acid, vanillin, vanillic acid and their applications
JOURNAL Patent: EP 0845532-A 29 03-JUN-1998;
HARMANN & REIMER GMBH (DE)
FEATURES
SOURCE location/Qualifiers
1. 1770
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
1. 1770
/gene="FCS"
1. 1770
/gene="FCS"
/codon_start=1
/product="FERULASABURE-COA-SYNTHEASE"
/protein_id="CAB69489.1"
/db_xref="GI:6740915"
/translation="MRSLEALPFPGRILERLEHMAKTRPEQTCVAAANGEMRRIS
LSQDFNVRAIAQSLPYGSAERPLIYSGNDLEHLOAFGMVAGIIPCPSPAYS
YADPFAKRIHYGLAOPGVAPADAPQRALETLPDVPALFTRGELAGRTVSF
DSLLEPGGIEADNAPATGPDITAKFLTSGSTKLPAKVPPTORMLCANQMLDTF
PVPGEPPVLYDWLPMNHFFGSHNIGIVYNGQTYVLDGKPTAGCPATLNNLSI
SPFAYVLPVGMELVGAERDSTLEREPFAKPLFPFAAGLSQSIWRLDVAROH
CGSRIMMAGLGMTEAPSCFTTGTSMAGYIGLPAPGCEVLIVDGDLEGRFHGP
HWSGYYRPAEONAAQFDESGYCSGDATKLADPADPOKGLMDGRIAEDEFKSSGVF
VSGPILRTAVILEGSSVYLDVVAADREGLGVLPRLDCAISGLGKEASDAVYL
ASPPVAPAFADIKRLNREITGNASIRIMVGLIDTPPSIDKGEVTKGSIINQAAVLQW
RSKAVVALYRGEDQSMRLDEATL"

ORIGIN

Query Match 100.0%; Score 1770; DB 6; Length 1770;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGCGTTCTCGAGGCGCTTTCTTCCCTTCCGGGTGGAATTCTTGAGCCTTCGAGCAT 60
DB 1 ATGCGTTCTCGAGGCGCTTTCTTCCCTTCCGGGTGGAATTCTTGAGCCTTCGAGCAT 60
QY 61 TGGGCTAAGACCCGTCAGAAACAACCTGCGTTGCTGCGAGGCGGCAATGGGGAATGG 120
DB 61 TGGGCTAAGACCCGTCAGAAACAACCTGCGTTGCTGCGAGGCGGCAATGGGGAATGG 120
QY 121 CGTGTATCAGTACCGGAAATGTTTCACACGTCGCGCATTCGACAGACTTGCTT 180
DB 121 CGTGTATCAGTACCGGAAATGTTTCACACGTCGCGCATTCGACAGACTTGCTT 180
QY 181 CCTTACGGAATATCGGAGAGCGCTCGCTTATCGTCTCTGGAATGACCTTGAAAT 240
DB 181 CCTTACGGAATATCGGAGAGCGCTCGCTTATCGTCTCTGGAATGACCTTGAAAT 240
QY 241 CTTACGCTGCAATTTGGGGCTATGATGCGGCAATTCCTATTGCGCGGTCTCTGCT 300
DB 241 CTTACGCTGCAATTTGGGGCTATGATGCGGCAATTCCTATTGCGCGGTCTCTGCT 300
QY 301 TATTCACTGCTGCGAAAGATTGGCGAGCTGCGTCAATCGTAGTCTTTCGAAACG 360
DB 301 TATTCACTGCTGCGAAAGATTGGCGAGCTGCGTCAATCGTAGTCTTTCGAAACG 360
QY 361 GGAATGCTCTTTCGCGCATTCGAGCATTTCGAGGCGCAATTGAGACATTTTCGCG 420
DB 361 GGAATGCTCTTTCGCGCATTCGAGCATTTCGAGGCGCAATTGAGACATTTTCGCG 420
QY 421 GAGAGCGTCCGCGCAATCTTCACTCGAGCGAAATTGGCGGCGCGCAAGTGAATTT 480
DB 421 GAGAGCGTCCGCGCAATCTTCACTCGAGCGAAATTGGCGGCGCGCAAGTGAATTT 480
QY 481 GACAGCTGCTGAGAGCAGCTGCTGGAGATTGAGGCAATATATCCCTTTCGCGCACT 540
DB 481 GACAGCTGCTGAGAGCAGCTGCTGGAGATTGAGGCAATATATCCCTTTCGCGCACT 540
QY 541 CCGGATACGATTCGCAAGTTCTTGTTCATTCTGCGCTTCAACAACTGCTTAAGCGGT 600
DB 541 CCGGATACGATTCGCAAGTTCTTGTTCATTCTGCGCTTCAACAACTGCTTAAGCGGT 600
QY 601 CCGACTACTGAGGGAATGCTCTGCGCAATCAGCAAGTCTTCTGCAAACTTTCCGCT 660
DB 601 CCGACTACTGAGGGAATGCTCTGCGCAATCAGCAAGTCTTCTGCAAACTTTCCGCT 660
QY 661 TTTGGTGAAGAGCCGCGGTGCTGTGTGAGCTGTGTCGCGTGAACCAACACTTCGCGGC 720
DB 661 TTTGGTGAAGAGCCGCGGTGCTGTGTGAGCTGTGTCGCGTGAACCAACACTTCGCGGC 720
QY 721 AGGCACAACATCGGCATCGTGTGTAACAAGCGGCGCACTGACCTTGAACGAGTAA 780
DB 721 AGGCACAACATCGGCATCGTGTGTAACAAGCGGCGCACTGACCTTGAACGAGTAA 780
QY 781 CCAACCGCCCAAGGTTTCGCGAGCGCTTCGCACTTGAAGCAAACTCTCCCACTGCG 840
DB 781 CCAACCGCCCAAGGTTTCGCGAGCGCTTCGCACTTGAAGCAAACTCTCCCACTGCG 840
QY 841 TACCTCACTGTGCGAAGAGCTGAGGAATTAATGTGGTGCCTTTGAGCAGACAGTACC 900
DB 841 TACCTCACTGTGCGAAGAGCTGAGGAATTAATGTGGTGCCTTTGAGCAGACAGTACC 900
QY 901 CTGGGGAAGAGCTTCTGCTGCGATGAAGCTTCTTCTTCCGCGGCGGTGGGTGTCG 960
DB 901 CTGGGGAAGAGCTTCTGCTGCGATGAAGCTTCTTCTTCCGCGGCGGTGGGTGTCG 960
QY 961 CAAGGATCTGGAGTCTTTTGAACCGGCTGCTGAACAGACTGTGGTGAACCATTCGC 1020
DB 961 CAAGGATCTGGAGTCTTTTGAACCGGCTGCTGAACAGACTGTGGTGAACCATTCGC 1020
QY 1021 ATGATGCGGGGTCTGGGCAATGACGGAAGTGTCTCTTCTGCACTTTTACACCGGACG 1080
  
```

```

DB 1021 ATGATGCGGGGTCTGGGCAATGACGGAAGTGTCTCTTCTGCACTTTTACACCGGACG 1080
QY 1081 CTGTGATGCGGTCTTGAACATTGGGCTGCGAAGCTGCGGAGTCAAGCTCGTTCCG 1140
DB 1081 CTGTGATGCGGTCTTGAACATTGGGCTGCGAAGCTGCGGAGTCAAGCTCGTTCCG 1140
QY 1141 GTGATGGAAGAAATTGGAAGGCGCTTTCATGTGTCGCAAGTCAATGAGCGGCTACTG 1200
DB 1141 GTGATGGAAGAAATTGGAAGGCGCTTTCATGTGTCGCAAGTCAATGAGCGGCTACTG 1200
QY 1201 GCTCTGAAACAAATGCCAAGCTTTCAGAGGAGGCTATTATCTGCTCCGCTGATGCC 1260
DB 1201 GCTCTGAAACAAATGCCAAGCTTTCAGAGGAGGCTATTATCTGCTCCGCTGATGCC 1260
QY 1261 ATCAAAATGGAAGATCTGCGCATCTCTCAAGAAAGCTGATGTTTGAACGCTGAATTGCT 1320
DB 1261 ATCAAAATGGAAGATCTGCGCATCTCTCAAGAAAGCTGATGTTTGAACGCTGAATTGCT 1320
QY 1321 GAAGACTTCAAGCTGCTCAGGGGTAATTTGTCAGCGTTGGGCAATTCGCAACGCGGCG 1380
DB 1321 GAAGACTTCAAGCTGCTCAGGGGTAATTTGTCAGCGTTGGGCAATTCGCAACGCGGCG 1380
QY 1381 GTTCTGGAAGGCGGCTCTTACGTTCTTGAACGTAATGTGCTGCTCTGATCTGAATGC 1440
DB 1381 GTTCTGGAAGGCGGCTCTTACGTTCTTGAACGTAATGTGCTGCTCTGATCTGAATGC 1440
QY 1441 CTTGATGCTGCTGTTTCCGCGCTTTCGAGCTGCGGCTTTCGAGGAGCTGAGAAA 1500
DB 1441 CTTGATGCTGCTGTTTCCGCGCTTTCGAGCTGCGGCTTTCGAGGAGCTGAGAAA 1500
QY 1501 GAGGCGTCCGAGCGCCAGAGTCTTTCGAGTACCGGCTTCGCGCTTCGAGGAGCTGAG 1560
DB 1501 GAGGCGTCCGAGCGCCAGAGTCTTTCGAGTACCGGCTTCGCGCTTCGAGGAGCTGAG 1560
QY 1561 CTCAAACTCAATGGAAGCAATGGAAGCAATGGAAGCAATGGAAGCAATGGAAGCAAT 1620
DB 1561 CTCAAACTCAATGGAAGCAATGGAAGCAATGGAAGCAATGGAAGCAATGGAAGCAAT 1620
QY 1621 CTGATACGCGCGCGTCAATGTAATGAGGCGAGGTCACTGACAAAGGCTGATCAACAG 1680
DB 1621 CTGATACGCGCGCGTCAATGTAATGAGGCGAGGTCACTGACAAAGGCTGATCAACAG 1680
QY 1681 CCGGCTGTTTTCGAAATGCGGTGCGGAAAGTTGATGCGCTGATGCTGTGAATCAA 1740
DB 1681 CCGGCTGTTTTCGAAATGCGGTGCGGAAAGTTGATGCGCTGATGCTGTGAATCAA 1740
QY 1741 TCCATGCTGCTGACGAGGCGCACTGTGA 1770
DB 1741 TCCATGCTGCTGACGAGGCGCACTGTGA 1770

```

RESULT 2
 PSP238746 4166 bp DNA linear BCT 05-NOV-1999
 LOCUS
 DEFINITION
 Pseudomonas sp. vdh (partial), fcs, aat and mac (partial) genes, strain HR199.
 ACCESSION
 AJ238746
 VERSION
 AJ238746.1 GI:6273620
 KEYWORDS
 aat gene; beta-ketothiolase; fcs gene; feruloyl-CoA synthetase; mac gene; methyl-accepting chemotaxis protein; vanillin dehydrogenase; vdh gene.
 SOURCE
 Pseudomonas sp.
 ORGANISM
 Pseudomonas sp.
 Bacteria; Proteobacteria.
 REFERENCE
 1 Overhage, J., Priefert, H. and Steinbuechel, A. Biochemical and genetic analyses of ferulic acid catabolism in Pseudomonas sp. Strain HR199 Appl. Environ. Microbiol. 65 (11), 4837-4847 (1999)
 JOURNAL
 MEDLINE
 PUBMED
 20011220
 10543794
 2 (bases 1 to 4166)
 REFERENCE

AUTHORS
TITLE
JOURNAL

Priefert H.
Submitted (04-MAY-1999) Priefert H., Westfälische
Wilhelms-Universität Münster, Institut für Mikrobiologie,
Corrensstrasse 3, Münster, D-48149, GERMANY

FEATURES

Location/Qualifiers

SOURCE

1. .4166

/organism="Pseudomonas sp."

/mol_type="genomic DNA"

/strain="HR199"

/db_xref="taxon:306"

/country="Indonesia"

1. .170

/gene="vdh"

<1. .128

/function="NAD dependent oxidation of vanillin to vanillic
acid"

/codon_start=3

/evidence=experimental

/transl_table=11

/product="vanillin dehydrogenase"

/protein_id="CAB60225.1"

/db_xref="GI:6273621"

/db_xref="GOA:O9RLD0"

/translation="QMFGGVKSSGYSFGSRASIEHTQLRLTONGRRHPI"

138. .170

/gene="vdh"

/function="putative terminator"

join(373. .378,388. .2157)

/gene="fca"

373. .378

/gene="fca"

388. .2157

/gene="fca"

/function="ATP dependent conversion of ferulic acid to
feruloyl-Coenzyme A"

/codon_start=1

/evidence=experimental

/transl_table=11

/product="feruloyl-CoA synthetase"

/protein_id="CAB60226.1"

/db_xref="GI:6273622"

/db_xref="GOA:O9RLD9"

/db_xref="TrEMBL:O9RLD9"

/translation="MRSLEALPPGRILRLHEMAKTRBEQTCVAAANGEMRRIS
LAEAFNVRAIAOSILPYGLSAERPLIVSGNDLEHLQAFAMVAGIPLYCPSPAYS
LSODLAKLTHIVGLQPLVFPADAPFORAETILPDVPAITFTGELAGRTVAF
DILLEOGGLEADNAPATGPTIAKELFTSGSKLPKAVPTTORMCANOMLQTF
PVGEPEPVLDMLPMHTTGGSHNIGIVLNGATYLLDGGKPTAGCFATRLRLSEI
SPFAYLVTPKGMELVGLERDSTLRERFPARMKLFPFAAGLSQGIWDRIDRAEBOH
GGRIRMMAGIEMTERPASCFTTGPISMGYIGLPAQGEVKLVQPDGLRGEFHBP
HVMGVRMAGIEMTERPASCFTTGPISMGYIGLPAQGEVKLVQPDGLRGEFHBP
VSSPRLTRAVLEGGSYVLDVVVAPADRECIQLVFPRLDCLRLSGIGBASAEVL
ASPPVRAWFPADMLKRLREATGNASRLMWGILDTTPSIDKGEVTDGKSNQRAVLQW
RSKAKVDALYRGEODSHLRDEATL"

2143. .2146

/gene="aat"

2143. .2146

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

FEATURES

Location/Qualifiers

SOURCE

1. .4166

/organism="Pseudomonas sp."

/mol_type="genomic DNA"

/strain="HR199"

/db_xref="taxon:306"

/country="Indonesia"

1. .170

/gene="vdh"

<1. .128

/function="NAD dependent oxidation of vanillin to vanillic
acid"

/codon_start=3

/evidence=experimental

/transl_table=11

/product="vanillin dehydrogenase"

/protein_id="CAB60225.1"

/db_xref="GI:6273621"

/db_xref="GOA:O9RLD0"

/translation="QMFGGVKSSGYSFGSRASIEHTQLRLTONGRRHPI"

138. .170

/gene="vdh"

/function="putative terminator"

join(373. .378,388. .2157)

/gene="fca"

373. .378

/gene="fca"

388. .2157

/gene="fca"

/function="ATP dependent conversion of ferulic acid to
feruloyl-Coenzyme A"

/codon_start=1

/evidence=experimental

/transl_table=11

/product="feruloyl-CoA synthetase"

/protein_id="CAB60226.1"

/db_xref="GI:6273622"

/db_xref="GOA:O9RLD9"

/db_xref="TrEMBL:O9RLD9"

/translation="MRSLEALPPGRILRLHEMAKTRBEQTCVAAANGEMRRIS
LAEAFNVRAIAOSILPYGLSAERPLIVSGNDLEHLQAFAMVAGIPLYCPSPAYS
LSODLAKLTHIVGLQPLVFPADAPFORAETILPDVPAITFTGELAGRTVAF
DILLEOGGLEADNAPATGPTIAKELFTSGSKLPKAVPTTORMCANOMLQTF
PVGEPEPVLDMLPMHTTGGSHNIGIVLNGATYLLDGGKPTAGCFATRLRLSEI
SPFAYLVTPKGMELVGLERDSTLRERFPARMKLFPFAAGLSQGIWDRIDRAEBOH
GGRIRMMAGIEMTERPASCFTTGPISMGYIGLPAQGEVKLVQPDGLRGEFHBP
HVMGVRMAGIEMTERPASCFTTGPISMGYIGLPAQGEVKLVQPDGLRGEFHBP
VSSPRLTRAVLEGGSYVLDVVVAPADRECIQLVFPRLDCLRLSGIGBASAEVL
ASPPVRAWFPADMLKRLREATGNASRLMWGILDTTPSIDKGEVTDGKSNQRAVLQW
RSKAKVDALYRGEODSHLRDEATL"

2143. .2146

/gene="aat"

2143. .2146

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

2154. .3449

/gene="aat"

ORIGIN

Query Match 100.0%; Score 1770; DB 1; Length 4166;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGCGTTCCTCGAGGCGCTTCTCCCTCCCGGTCGAATTTCTTGAAGCGTTCGAGCAT 60
388 ATGCGTTCCTCGAGGCGCTTCTCCCTCCCGGTCGAATTTCTTGAAGCGTTCGAGCAT 447
61 TGGGCTTAAGCCCGTCCGAAACAACCTGGCTTCTCCGAGGCGGCAATGGGGAATGG 120
448 TGGGCTTAAGCCCGTCCGAAACAACCTGGCTTCTCCGAGGCGGCAATGGGGAATGG 507
121 CGTGTATCAGCTACGCGGGAATGTTCCACAACGTCGCGCCATCGACAGAGCTTCTT 180
508 CGTGTATCAGCTACGCGGGAATGTTCCACAACGTCGCGCCATCGACAGAGCTTCTT 567
181 CTTTACGAGTATCGGAGAGCGTCCGCTTATCGTCTCTGGAATGACCTTGAACAT 240
568 CTTTACGAGTATCGGAGAGCGTCCGCTTATCGTCTCTGGAATGACCTTGAACAT 627
241 CTTTACGAGTATCGGAGAGCGTCCGCTTATCGTCTCTGGAATGACCTTGAACAT 300
628 CTTTACGAGTATCGGAGAGCGTCCGCTTATCGTCTCTGGAATGACCTTGAACAT 687
301 TATTCACTGCTGTCGCAAGATTTGGCGAAGTGGCTGACATCGTAGGCTTTCTGCAACG 360
688 TATTCACTGCTGTCGCAAGATTTGGCGAAGTGGCTGACATCGTAGGCTTTCTGCAACG 747
361 GGAATGCTCTTGTGTCGCAAGTGGCTGACATCGTAGGCTTTCTGCAACG 420
748 GGAATGCTCTTGTGTCGCAAGTGGCTGACATCGTAGGCTTTCTGCAACG 807
421 GAGCAGCTGCGCCGCAATCTTCACTCGAGGCGAATTTGGCGGCGGCGCAAGTGAATTT 480
808 GAGCAGCTGCGCCGCAATCTTCACTCGAGGCGAATTTGGCGGCGGCGCAAGTGAATTT 867
481 GAGCAGCTGCGCCGCAATCTTCACTCGAGGCGAATTTGGCGGCGGCGCAAGTGAATTT 940
868 GAGCAGCTGCGCCGCAATCTTCACTCGAGGCGAATTTGGCGGCGGCGCAAGTGAATTT 927
941 GAGCAGCTGCGCCGCAATCTTCACTCGAGGCGAATTTGGCGGCGGCGCAAGTGAATTT 1000
928 GAGCAGCTGCGCCGCAATCTTCACTCGAGGCGAATTTGGCGGCGGCGCAAGTGAATTT 987
601 CCGACTACTAGGGAAGTCTTGGCGCAATCTTCACTCGAGGCGAATTTGGCGGCGGCG 660
988 CCGACTACTAGGGAAGTCTTGGCGCAATCTTCACTCGAGGCGAATTTGGCGGCGGCG 1047
661 TTGGTGAAGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720

```
Db 1048 TTTGTGAAGAGCGCGGTGCTGTGGAATGCTGTGCGGAAACACACTTCGCGGCG
Qy 721 AGCCACAACATCGGCATCGTGTGTGTAACAAGCGGCGACGTAACCTTGACGACGGTAA 780
Db 1108 AGCCACAACATCGGCATCGTGTGTGTAACAAGCGGCGACGTAACCTTGACGACGGTAA 1167
Qy 781 CCAACCGCCCAAGGGTTCGCGGACGCTTCGCACTTGAGCGAAATCTCTCCCACTGCG 840
Db 1168 CCAACCGCCCAAGGGTTCGCGGACGCTTCGCACTTGAGCGAAATCTCTCCCACTGCG 1227
Qy 841 TACCTCACTGTGCGGAAAGGCTGAGAGAAATTAATGTGGTGCCCTTGAGCGAGACAGTACC 900
Db 1228 TACCTCACTGTGCGGAAAGGCTGAGAGAAATTAATGTGGTGCCCTTGAGCGAGACAGTACC 1287
Qy 901 CTGGCGGAAGCGCTTCTTCGCTGCAATGAACTGTCTCTTCGCGGCGGTGGGTTGTCG 960
Db 1288 CTGGCGGAAGCGCTTCTTCGCTGCAATGAACTGTCTCTTCGCGGCGGTGGGTTGTCG 1347
Qy 961 CAAGGATCTGGAGTGTGGAACCGGATCGTGAAACAGACTGTGGTGAAGCCATTGCG 1020
Db 1348 CAAGGATCTGGAGTGTGGAACCGGATCGTGAAACAGACTGTGGTGAAGCCATTGCG 1407
Qy 1021 ATGATGCGGCGGTCTGGGATGACGAGAGCTGTCTCTTCTGCACTTTTACCCGAGACG 1080
Db 1408 ATGATGCGGCGGTCTGGGATGACGAGAGCTGTCTCTTCTGCACTTTTACCCGAGACG 1467
Qy 1081 CTGTCGATGCGCTGTTACATTTGGGCTGCGAGCGCGCTGGGCGAGAGCTCGATTGCG 1140
Db 1468 CTGTCGATGCGCTGTTACATTTGGGCTGCGAGCGCGCTGGGCGAGAGCTCGATTGCG 1527
Qy 1141 GTGATGGAATAATGGAAGGCGCTTTCATGATCGGACGCTCATGAGCGGCTACTGCGCT 1200
Db 1528 GTGATGGAATAATGGAAGGCGCTTTCATGATCGGACGCTCATGAGCGGCTACTGCGCT 1587
Qy 1201 GCTCTGGAATAATGCGCAAGCTTTCGAGAGAGGCTTATTACTGCTCCGCTGATGCC 1260
Db 1588 GCTCTGGAATAATGCGCAAGCTTTCGAGAGAGGCTTATTACTGCTCCGCTGATGCC 1647
Qy 1261 ATCAAAATGGGCAATCTGCGGATCTCGAGAGGCTGATGTTGACGCGTGCAGATTGCT 1320
Db 1648 ATCAAAATGGGCAATCTGCGGATCTCGAGAGGCTGATGTTGACGCGTGCAGATTGCT 1707
Qy 1321 GAAGACTTCAAGCTGTCTTCAGGAGGTAATTTGTACGCTGTGGCCATTTGCGACGCGGCG 1380
Db 1708 GAAGACTTCAAGCTGTCTTCAGGAGGTAATTTGTACGCTGTGGCCATTTGCGACGCGGCG 1767
Qy 1381 GTTCTGGAAGGCGGCTTTCAGCTGCGAGAGTGTGCTGCTCTGATCGTGAATGC 1440
Db 1768 GTTCTGGAAGGCGGCTTTCAGCTGCGAGAGTGTGCTGCTCTGATCGTGAATGC 1827
Qy 1441 CTGGAATGCTGCTGTTCCGCGCTTCTGACATGCGCGGCTTGTGCGGAGCTAGGAAA 1500
Db 1828 CTGGAATGCTGCTGTTCCGCGCTTCTGACATGCGCGGCTTGTGCGGAGCTAGGAAA 1887
Qy 1501 GAGGCGTCCGACCGCGAGGTCTTTCAGAGTGAAGCGCGGTTGCGGCTGTGCTGACTGCG 1560
Db 1888 GAGGCGTCCGACCGCGAGGTCTTTCAGAGTGAAGCGCGGTTGCGGCTGTGCTGACTGCG 1947
Qy 1561 CTCAAACTGATGATGAGAGAGCACTGCGAATGCCAGTGTGATCATGTGGGTAGGGCTC 1620
Db 1948 CTCAAACTGATGATGAGAGAGCACTGCGAATGCCAGTGTGATCATGTGGGTAGGGCTC 2007
Qy 1621 CTGATACGCGCGCGCTGATGATTAAGGCGAGAGTCACTGACAAAGGCTCGATCAACGAG 1680
Db 2008 CTGATACGCGCGCGCTGATGATTAAGGCGAGAGTCACTGACAAAGGCTCGATCAACGAG 2067
Qy 1681 CGGCTGTTTGGCAATGCGCGGTGCGCAAAAGTGTGATCGCTGTATCGTGTGAAGTCAA 1740
Db 2068 CGGCTGTTTGGCAATGCGCGGTGCGCAAAAGTGTGATCGCTGTATCGTGTGAAGTCAA 2127
Qy 1741 TCCATGCTGCTGATGAGAGGCAACACTGTGTA 1770
```

```
Db 2128 TCCATGCTGCTGATGAGAGGCAACACTGTGTA 2157
RESULT 3
A92090
LOCUS A92090 32679 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0845532.
ACCESSION A92090
VERSION A92090.1 GI:6740886
KEYWORDS
SOURCE Pseudomonas sp.
ORGANISM Pseudomonas sp.
Bacteria; Proteobacteria.
REFERENCE 1 (bases 1 to 32679)
AUTHORS Priefer,H.D. and Rabenhorst,J.D.
TITLES Enzymes for the synthesis of coniferyl alcohol, coniferyl aldehyde,
ferulic acid, vanillin, vanillic acid and their applications
JOURNAL Patent: EP 0845532-A 1 03-JUN-1998;
HARMANN & REIMER GMBH (DE)
FEATURES
source
1..32679
/organism="Pseudomonas sp."
/mol_type="unassigned DNA"
/strain="HR19"
/db_xref="taxon:306"
3146..4000
/gene="ORF1"
3146..4000
/gene="ORF1"
/codon_start=1
/strain_table=1
/protein_id="CAB69475.1"
/db_xref="GI:6740887"
/translation="MIATGAGAGGCGCTTCCCTTCCGCGGTGCAATTTCTGAGGCTCGAGCAT
GIVRQADYNREPETHRALIGVNRLLISSVGGRTAQRPAVIDAKXEGEILAT
SLHADKSLGATEHRTDQLTSGITPHVLRNMTHTENTAGIPVALVHCVLLG
ADGGILASARADVAEAAVVLTEGQACRVLEGPATVITELAAVAPQKTV
YENLESDDYRSALISNGLPDPFALLADSDAGAAKGYLPDSSGDSRKLGRPTTMS
AIAAAGR"
ORIGIN
Query Match 100.0%; Score 1770; DB 6; Length 32679;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ANGCGTCTCTCGAGGCGCTTCCCTTCCGCGGTGCAATTTCTGAGGCTCGAGCAT 60
Db 2128 ATGCGTCTCTCGAGGCGCTTCCCTTCCGCGGTGCAATTTCTGAGGCTCGAGCAT 21347
Qy 61 TGGGCTAAGACCCGTCAGAAACAACCTGCTGTGCGAGGCGGCAATGGGAATGG 120
Db 21348 TGGGCTAAGACCCGTCAGAAACAACCTGCTGTGCGAGGCGGCAATGGGAATGG 21407
Qy 121 CTTGATACAGCTACCGGAAATGTTCCACAACTCCGCGCATCGCACAGACTTGCTT 180
Db 21408 CTTGATACAGCTACCGGAAATGTTCCACAACTCCGCGCATCGCACAGACTTGCTT 21467
Qy 181 CTTTACGACCTATCGGACAGAGCGTCCGCTTATCGTCTCTGGAATAGACCTGAAACAT 240
Db 21468 CTTTACGACCTATCGGACAGAGCGTCCGCTTATCGTCTCTGGAATAGACCTGAAACAT 21527
Qy 241 CTTACGCTGCAATTTGGGCTATGTAATGCGGAGATTCCCTATTGCCGCTGCTCCTGCT 300
Db 21528 CTTACGCTGCAATTTGGGCTATGTAATGCGGAGATTCCCTATTGCCGCTGCTCCTGCT 21587
Qy 301 TATTCACTGCTGTGCGAAGATTTGGGAGAGCTGCGTCAATGTAAGTCTTCTGCAACCG 360
Db 21588 TATTCACTGCTGTGCGAAGATTTGGGAGAGCTGCGTCAATGTAAGTCTTCTGCAACCG 21647
Qy 361 GGAATGCTTTGCTGCGATGAGACACTTTCACAGCGGCAATTGAGACATTTGCGCG 420
Db 21648 GGAATGCTTTGCTGCGATGAGACACTTTCACAGCGGCAATTGAGACATTTGCGCG 21707
```

QY 421 GAGCAGCTGCCCCGCAATCTTCACTCGAGCGCAATTGGCCGGGCGGCGACGCGTGTATTT 480
 DB 21708 GAGCAGCTGCCCCGCAATCTTCACTCGAGCGCAATTGGCCGGGCGGCGACGCGTGTATTT 21767
 QY 481 GAGCAGCTGCTGAGCAGCCTGTGTGGATTGAGGCAATATATGCTTTTGGCAACTGAC 540
 DB 21768 GAGCAGCTGCTGAGCAGCCTGTGTGGATTGAGGCAATATATGCTTTTGGCGCAACTGAC 21827
 QY 541 CCGGATACGATTGGCAAGTTCTTGTCACTTGTGCTTACCAAACTGCTTAAGCCGTG 600
 DB 21828 CCGGATACGATTGGCAAGTTCTTGTCACTTGTGCTTACCAAACTGCTTAAGCCGTG 21887
 QY 601 CCGGATACGATTGGCAAGTTCTTGTGCGCAATGAGATGCTTTCGAACTTTCGCGGT 660
 DB 21888 CCGGATACGATTGGCAAGTTCTTGTGCGCAATGAGATGCTTTCGAACTTTCGCGGT 21947
 QY 661 TTTGGTGAAGAGCCGCGGTGCTGGTGAAGCTTGGCGTGAACAACCTTTCGCGGC 720
 DB 21948 TTTGGTGAAGAGCCGCGGTGCTGGTGAAGCTTGGCGTGAACAACCTTTCGCGGC 22007
 QY 721 AGCCCAACAATCCGATCTGTGTGTACACGCGCGCACTGTAACCTTGAACGCGTAA 780
 DB 22008 AGCCCAACAATCCGATCTGTGTGTACACGCGCGCACTGTAACCTTGAACGCGTAA 22067
 QY 781 CCAACCGCCCAAGGGTTCGCGGAGAGCGCTTGCACATTTGAGCGAAATCTTCCACCTGCG 840
 DB 22068 CCAACCGCCCAAGGGTTCGCGGAGAGCGCTTGCACATTTGAGCGAAATCTTCCACCTGCG 22127
 QY 841 TACCTCACTGTGCGAAAGCTGGAGGAAATTAGTGGGTGCTTGAACGAGACAGTACC 900
 DB 22128 TACCTCACTGTGCGAAAGCTGGAGGAAATTAGTGGGTGCTTGAACGAGACAGTACC 22187
 QY 901 CTGCGCGCAAGCTTCTTCTGCTCGCATGAAGCTTCTTCTGCGCGCGCTGCGTGTG 960
 DB 22188 CTGCGCGCAAGCTTCTTCTGCTCGCATGAAGCTTCTTCTGCGCGCGCTGCGTGTG 22247
 QY 961 CAAGGAGCTGGGATCGTTTGGACCGGGTGCCTGAACAGACAGTGGTGAAGCGCATTCG 1020
 DB 22248 CAAGGAGCTGGGATCGTTTGGACCGGGTGCCTGAACAGACAGTGGTGAAGCGCATTCG 22307
 QY 1021 ATGATGCGCGGATCTGGGATGACGAGACAGTCTCTTCTGCACTTTTACACCGGACG 1080
 DB 22308 ATGATGCGCGGATCTGGGATGACGAGACAGTCTCTTCTGCACTTTTACACCGGACG 22367
 QY 1081 CTGTGATGCTGTGATCATTTGGGCTGCCAGCGCTGCTGCGAGGTCAAGCTGTTCCG 1140
 DB 22368 CTGTGATGCTGTGATCATTTGGGCTGCCAGCGCTGCTGCGAGGTCAAGCTGTTCCG 22427
 QY 1141 GTGATGCGGAAATTTGAAAGGCGTTTCCATGCTCCGACGTCATGAGCGGCTACTGGCGT 1200
 DB 22428 GTGATGCGGAAATTTGAAAGGCGTTTCCATGCTCCGACGTCATGAGCGGCTACTGGCGT 22487
 QY 1201 GCTCTGGAACAAATGCGCAAGCTTTCGAGAGGAGGCTATTACTGCTCGGTGATGCC 1260
 DB 22488 GCTCTGGAACAAATGCGCAAGCTTTCGAGAGGAGGCTATTACTGCTCGGTGATGCC 22547
 QY 1261 ATCAAAATGGCAATCTCTCCGATCTCAAGAAAGTCTGATGTTTGAAGCTCGAATGCT 1320
 DB 22548 ATCAAAATGGCAATCTCTCCGATCTCAAGAAAGTCTGATGTTTGAAGCTCGAATGCT 22607
 QY 1321 GAAGACCTTCAAGCTCTCAGGGGTATTGTCAAGCGTTGGCCATTGGCCACGCGGGCG 1380
 DB 22608 GAAGACCTTCAAGCTCTCAGGGGTATTGTCAAGCGTTGGCCATTGGCCACGCGGGCG 22667
 QY 1381 GTTCTGGAAGCGGCTTCTAAGCTCTGAGAGTATGTTGCTGCTCTGATTCGTAATGC 1440
 DB 22668 GTTCTGGAAGCGGCTTCTAAGCTCTGAGAGTATGTTGCTGCTCTGATTCGTAATGC 22727
 QY 1441 GTTGAATGCTCTGTTTCCGCGCTTCTCTGACTGCGCGCTTGTGCGGAGTAAAGAAA 1500
 DB 22728 GTTGAATGCTCTGTTTCCGCGCTTCTCTGACTGCGCGCTTGTGCGGAGTAAAGAAA 22787
 QY 1501 GAGGCGTCGAGCGCGAGGTGCTTGCAGATGAGCCGCTGCGGCTGTGTTGCTGACTGG 1560

DB 22788 GAGGCGTCGAGCGCCGAGAGCTGCTTGCAGTGAACCCGCTTGGGCGCTGTTGCTGACTGG 22847
 QY 1561 CTCAAACGACTCAATCGAAGAGCACTGCGCAATGCCAGTGCATCATGTGGTAAAGGCTC 1620
 DB 22848 CTCAAACGACTCAATCGAAGAGCACTGCGCAATGCCAGTGCATCATGTGGTAAAGGCTC 22907
 QY 1621 CTCGATACGCGCGCTCGATTTGATTAAGGAGAGTCACTGACAAAGGCGCTGATCAACAG 1680
 DB 22908 CTCGATACGCGCGCTCGATTTGATTAAGGAGAGTCACTGACAAAGGCGCTGATCAACAG 22967
 QY 1681 CGCGCTGTTTGGCAATGCGCGCTGCGCAAGTGTGATCGCTGATCGTGTGAAGATCAA 1740
 DB 22968 CGCGCTGTTTGGCAATGCGCGCTGCGCAAGTGTGATCGCTGATCGTGTGAAGATCAA 23027
 QY 1741 TCCATGCTGCGTGAAGAGGCGCACACTGTGA 1770
 DB 23028 TCCATGCTGCGTGAAGAGGCGCACACTGTGA 23057

RESULT 4
 PFL536325
 LOCUS
 DEFINITION
 Pseudomonas fluorescens ferr gene, ech gene, vdh gene, fcs gene,
 aat gene and fcd gene.
 ACCESSION
 AJ536325
 VERSION
 AJ536325.1 GI:40788116
 KEYWORDS
 aat gene; beta-ketothiolase; ech gene; fcd gene; fcs gene; ferr
 gene; feruloyl-CoA dehydrogenase; feruloyl-CoA hydratase;
 feruloyl-CoA synthetase; transcriptional regulator; vanillin
 dehydrogenase; vdh gene.
 SOURCE
 Pseudomonas fluorescens
 Pseudomonas fluorescens
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadaceae;
 Pseudomonadaceae; Pseudomonas.
 ORGANISM
 1
 Barhlini, P., Ciavolani, C., Picca, A. G., Schiesse, A. and Ruzzi, M.
 Cloning and Characterization of Ferulate Catabolism Genes from
 Pseudomonas fluorescens Bfl3
 Unpublished
 2 (bases 1 to 9477)
 REFERENCE
 Ruzzi, M.
 Direct Submission
 Submitted (09-JAN-2003) Ruzzi M., Agrobiology and Agrochemistry,
 DNBAC - University of Tuscia, via C. de Dellis, snc, Viterbo -
 01100, ITALY
 JOURNAL
 FEATURES
 source
 1..9477
 /organism="Pseudomonas fluorescens"
 /mol_type="genomic DNA"
 /strain="Bfl3"
 /isolation_source="soil"
 /db_xref="taxon:294"
 /environmental_sample
 /country="Italy:Lazio"
 /complement(1297..1785)
 /gene="ferr"
 /complement(1297..1785)
 /gene="ferr"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="transcriptional regulator"
 /protein_id="CAD60265.1"
 /db_xref="GI:40788117"
 /translation="MAKPAFLADPREAPPAEVOAFLPSALDELIGYAMRAQLKLF
 ONLIGLSADHLPAPQFASALATIDSNPGMOWDIALAPALAIPEQVAPLUNKJESSRLA
 VARCKRPDKRSIGIFLSTKGTILKELKOIAADSDVATSAESAERBELRLIKKYV
 QD"
 1994..2824
 /gene="ech"
 1994..2824
 /gene="ech"
 /gene="ech"

```
/codon_start=-1
/evidence=experimental
/translation_table=11
/product="feruloyl-CoA hydratase"
/protein_id="CAD60266.1"
/db_xref="GI:40788118"
/translation="MSNIEGRWTTVEIEDGIAMVILNREPKRNASPTLNEMIDV
LETLBODPAAGVILVTGAEAMTAGMDLKEFREVADGEIIOEKIRREASQWTL
RMARPTLIAMVMCMFCGGFSPVACDLAI CADEATFGEIIMWGI.PPGNLVSKAAD
TYVROSLEYIMTGTEFGGOKAEMGIIVDSVPLARREYITELARNILKEKNVIVRA
AKHGFRCRELTMEQNEVDIYATLDSRLDTGREGGKKQLDDKS1KPGIQAIXR
"
2921. .4369
/gene="vch"
/2921. .4369
/gene="vch"
/codon_start=-1
/evidence=experimental
/translation_table=11
/product="vanillin dehydrogenase"
/protein_id="CAD60267.1"
/db_xref="GI:40788119"
/translation="MLDVLILIGGOSRPARDETLERRPVTGAVSVAAATLDDAD
AAYAAQAAPFAAALAPNRRSRRLKAHQLOARSGEPTAAGETGAMAMTGNVR
LAANMLREASMTQINGEVTISDVPGSEFAMALRQCGVVLGIAPNWAPVITRAIA
MLACGNTVYLKASELSAVHRLIGVLDAGAGDVVNVISNAPWAAIVERTIAN
PAVRVNTGSTHVRIGVLELAHRLKPALEIGKAPLILDDADLDAVAARAGA
YFNOGQICMSTERLIYDAKVAADFVAKAVALTRAGDPSADSVIGSLIVSAGTR
IKALIDALSKGARIVVGGQLDGSIILOPTLDGVTSMRLYRESEFPAVILIRGSD
BEILRLANSEBFLSNAISRTGRALALAQVESGICHTNGTVDDEAMPFGVKS
SGISFGGKASIHFTQLRWVTLQHGPRHPI"
4628. .6397
/gene="fcs"
/4628. .6397
/gene="fcs"
/codon_start=-1
/translation_table=11
/product="feruloyl-CoA synthetase"
/protein_id="CAD60268.1"
/db_xref="GI:40788120"
/translation="MRSILKPIALPRLILDLVHMAVRPEQFIAREAGDMRLVS
YQMDSVRAIAGILRYGSLADKPAALISGNDIEHLQALAMVAGTIPCEPSAYS
LISQDPKLRHVCDLQPGILFVSDASATQRAIDAVLPETPLISVGGVPGKQSF
ASILAPPGKAADAAFAATGPDSIAKFLFTSGSTKLPKVITTTQRMICANQOMLQTF
PVFGEPPVYLVDMLPMNHTFGGSHNVGILVINGSTFLDGEKPTAGFAETLKEI
SPFAYLTVPKMELELVNLEODAEIARECFKRMSLFFPAAAGLSQSVMDRLDVAOH
CGRRITMMAGIGCTEAPSGCTTTTGPISMAGYIGLPAPGCVTLVVDGKLBGRFPG
HIMPGYRAAQCTAYFPDAGPTCGSDALKADADPQGLMDGRIADPFKISGTV
VSVGLNRFAVLEGAIPYVDLVVAAPDRECLALVPRLYECRLALGLAESEAVL
ASAPVQWFGMDLQCLNREASGNASRLLEWIALQVEPASIDRBEITDKGINRAVILQW
RAEQVENLYRGRPSTILRAEPK"
6394. .7641
/gene="aat"
/6394. .7641
/gene="aat"
/codon_start=1
/translation_table=11
/product="putative beta-ketochiolase"
/protein_id="CAD60269.1"
/db_xref="GI:40788121"
/translation="MSGAGQYSGFADVAIIEAVRTWVDLGDIAEISPIDGIKVR
EYLARAGIBQAVDSVLASMAQSPDAILPRHIGLYSGVGAVPALVQORICATGF
BILRQABOILRGVQLALCVASRSMRNPILAAITHRGFRIGAAVQFOMLALYDP
ALGLDMITADNLARRYGLTREAIDGVAICSHORALAAOQAFPAIEIVASQOREL
DVGQAGITRLPGRVERVSDSHRPSERDILARITIHGQVOTAGNSCAVVDGAAL
LVGRASACDRPLALIRASAVYGVAREPNGIGAPAIRILLQKMEBAEQIRSAEIT
NKAQSOVLAAPELIGDRLNVQGSIALGHPILASGILRLITLARKORENNIKYG
IAAACVGGQAGMALIENPAWRA"
7701. .9455
/gene="fcd"
/7701. .9455
/gene="fcd"
/codon_start=-1
```

```
/translation_table=11
/product="putative feruloyl-CoA dehydrogenase"
/protein_id="CAD60270.1"
/db_xref="GI:40788122"
/translation="MMSYQAPLRDMQFVLEHMLAEAPMRSPAFALDPLALQVLE
QARFSQGVLAFLANSGRQGRMADQVSTREBPADATRAYTEGGWPAALCEBALGG
QGLPOLIDALDEMLIYASNMAWAMPTGIAHAYLCLKTGAEMLOERYIPALISGASL
PTMCLTEPLIASDVLGILRCRABPDGSRIGSGSKLFIISGEBDLTDNLIHLVLRPL
DABEGRGISLEFVPRLEQDONGVRCNGIEHKMGISKSATCSLVFDAQGMVSEA
NMGIAAPFMMSNSARLHVLOGIGHYEAAMONALAYAGROOKRALRPAAGAAQAD
PIRYHPAMRVLLEIATSEGRALIGYMAAHLIDQEHHPLEAPROALQDLPI
IKAFTEQGFROASNALQVFGGYVAEPAIREITLDSIAMIYESNEI QANDILLR
KVLGDGGFALLLEQRDEABQGAAGVACAGPAGLADLAELQGLVGVQWALSD
ABPYPAADPFLALCGLSLALAMARAARAVSRNLPBGDPLRGAKTATAGFFSYLLPQ
VEQRMATVBSAQALALAFI"
ORIGIN
```

Query Match 55.8%; Score 987.6; DB 1; Length 9477;
Best Local Similarity 72.4%; Pred. No. 2,4e-226;
Matches 1281; Conservative 0; Mismatches 489; Indels 0; Gaps 0;

QY	1	ATGCGTTCCTCGAGGCGCTTCCCTTCCCGGCGCAATTCTTGAGGCTCGAGCAT	60
DB	4628	ATGCGCTCGTGAAGCGCTTGCGGCTTGCCGATCGCTTGCATCGCTGTGTCAC	4687
QY	61	TGGGCTAAGACCCGTCAGAAACAACCTCGTTGTGCGAGGGCGCAATGGGGAATGG	120
DB	4688	TGGGCGCGGTGGCGCCCGAGAGACCTTCATTTGCGCCGCGAGGCGCGGTGGCATGG	4747
QY	121	CGTCGTATCAGCTACGCGGAAATGTTCCACAACTCCGCGCCATCGGCACAGACTTGCTT	180
DB	4748	CGCTTGTAGCGCTATGAGGAGATGTCGACAGGTGCGGCAATCGCCAGGGCGCTGCTT	4807
QY	181	CCTTACGACATACGCGCAGAGCGCTCCGCTGCTTATCGCTCTGGAATAGCACTGAAACAT	240
DB	4808	CGTTACGCGCTGTGCGCCAGCAAGCGCGCTGCGCTCTCTCGGCGAATGACATCGAACAC	4867
QY	241	CTTCAGCTGGCAATTTGGGGCTATGTAATGCGGGCAATTCCTATTTCCCGGTGCTCTGCTP	300
DB	4868	CTGCAACTGGCCCTGGGGCGGATGTACGCGGGATTCCTGATTTGCCGGGTGCGGGCC	4927
QY	301	TATTCATCTGCTGCGAAGATTTGGGAGAGCTGCGTCAATCGTAGCTTCTGCAACCG	360
DB	4928	TATTCCTGCTGCTGCGAGACCTTTGCAAGCTGCGGCAATGTCGCACTGTTGACGCGG	4987
QY	361	GGACTGCTCTTTCGTCGCGATGCGACACCTTTCACAGCGCAATTAGACCAATTTGCGCG	420
DB	4988	GGGCTGCTGTTGCTACGCAACGCGACGCGCTTACACGCGGCGCATGCAAGCGGTGCTGCCG	5047
QY	421	GACGACGTCCCGCAATCTTCACTCGAGGCGAATTTGGCGGGCGGCGACGCTGAGTTT	480
DB	5048	CCCGAAGCGCGGTATGACGCGTGGCGTGGCCAGGTGCCGGGGCGGCGCAAGGCACTTC	5107
QY	481	GACAGCTCTGTGAGAGCGCTGTGGGATTTGAGGCGAATTAATGCTTTTGGCGAATGCGC	540
DB	5108	GCGAGCTCTGTGCGCGAGCGCGGTGGCGCGAGGCGGAATGCGGCTTTTCCGCAACGCGC	5167
QY	541	CCGATACGATTGCCAAGTTCTTGTTCACCTTGTGCTTACCAAACTGCTAAGCGGTG	600
DB	5168	CCGAGACGATGCGCAAGTTTCTTTTCACTTGGGCTTCCACAGCTGCCCAAGGCGGTG	5227
QY	601	CCGACTACTCAGCGAATGCTTGCAGCGCAATGACAGATGCTTCTGCAACTTTCCGGTT	660
DB	5228	ATCACACACCGAGCATGCTTGGCGCAACAGCAATGCTGCTGAGACCTTTCCGGTGG	5287
QY	661	TTTGGTGAAGAGCGCGGCTGTGTGTGACTGCTGTTCCGTGGAAACCAACTTTCCGGCGC	720
DB	5288	TTTCGGTGAAGAGCGCGGCTGTGTGTGACTGCTGTTCCGTGGAAACCAACTTTCCGGCGC	5347
QY	721	AGCCACAACATGCGGATGCTGTTGTACAACGCGGCGGACGTAACCTTGAACGAGGTAA	780
DB	5348	AGCCATTAACGTGGGATGCTGTTGTACAACGCGGCGGTAACCTTTTACCTGGAACGAGGCAAG	5407


```

Oy 781 CCAACCGCCCAAGGGTTCGCGAGACGCTTCGCAATTGACGAAATCTCTCCACTGCG 840
Db 5408 CCAACCGCCCAAGGGTTCGCTGAAACCTCTGCAACTCAAGAGATTTCCCAACGCC 5467
Oy 841 TACCTCACTGTGCGCAAGAGCGTGGAGAAATGATGGGTCCTTGGACGAGACATACC 900
Db 5468 TACCTCACTGTGCGCAAGAGCGTGGAGAGAACTGATCAACCCCTCGAGCAGACGCCAG 5527
Oy 901 CTGCGGGAAGCGTTCTTCGCTCGCATGAACTGTTCTTTCGCGGCGCTGGGTTCTCG 960
Db 5528 CTGCGGGAAGGTTTTCGAGCCCATGAGCTGTTTTCGCGCGCGCTGGCTTTTCG 5587
Oy 961 CAAGGATCTGGATCGTTTGGACCGGGTCTGTAAGACAGACTGTGTGAGCGCATTCGC 1020
Db 5588 CAAGGATCTGGATCGCTTGGACCGGGTCTGTAAGACAGACTGTGTGAGCGCATTCGC 5647
Oy 1021 ATGATGCGGGTCTGGCATGACGAGACCTGCTTTCCTGCACTTTTACACCGACCG 1080
Db 5648 ATGATGCGGGTCTGGCATGACGAGACCGCGCTTTCCTGCACTTTTACACCGACCG 5707
Oy 1081 CTGTGATGCTGCTTCAATTTGGGCTGCCAGCGCTGGCTGGAGGTCAAGCTCTTCG 1140
Db 5708 CTGTGATGCTGCTTCAATTTGGGCTGCCAGCGCTGGCTGGAGGTCAAGCTCTTCG 5767
Oy 1141 GTGATGCGGAAATTTGAAAGGCGCTTTCATGTCCTGCACTGTAAGCGGCTACTGCGT 1200
Db 5768 GTGATGCGGAAATTTGAAAGGCGCTTTCATGTCCTGCACTGTAAGCGGCTACTGCGT 5827
Oy 1201 GCTCTGAAACAAATGCGCCAAAGCGTTTGCAGAGAAAGCTATTATCTGCTCGGTATGCC 1260
Db 5828 GCGGCGCAGACAGCGCGCGAGGTGTTGACGCCCAAGGCTTTCACTGCTCGGGGATGCG 5887
Oy 1261 ATGAAATTTGGCAATCTGCTCCGATCTTCAAGAAAGTGTATGTTTGAACGCTCAATTGCT 1320
Db 5888 CTCAACCTGCGCCAGCGCGCGCAACCGCAACTGGGCTGTATGTTTGAACGCTCAATTGCT 5947
Oy 1321 GAAGACTTCAAGCTGCTCTCAGAGGGTATTGTGACGCTGTGGGCAATGCGCAGCGCGCG 1380
Db 5948 GAAGACTTCAAGCTGCTCTCAGAGGGTATTGTGACGCTGTGGGCAATGCGCAGCGCGCG 6007
Oy 1381 GTTCTGAAAGCGCGCTTCTTACGTCCTGAGCGTATGCTGCTGCTGCTGCTGCTGCTGCTG 1440
Db 6008 GTTCTGAAAGCGCGCTTCTTACGTCCTGAGCGTATGCTGCTGCTGCTGCTGCTGCTGCTG 6067
Oy 1441 CTGGAATGCTGCTGCTTTCGCGCTCTTCTGACCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 1500
Db 6068 CTGGAATGCTGCTGCTTTCGCGCTCTTCTGACCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 6127
Oy 1501 GAGGCGTCCGACCGCGAGGTGCTTGCAGTGAAGCGGCTTGGGCTGCTGCTGCTGCTGCTGCTG 1560
Db 6128 GAGGCGTCCGACCGCGAGGTGCTTGCAGTGAAGCGGCTTGGGCTGCTGCTGCTGCTGCTGCTG 6187
Oy 1561 CTCAAAAGACTCAATGAGAGAACTGCGCAATGCGCATGATGAGGCTC 1620
Db 6188 TTGCAAGTCTGATGCTGAGGCGCAAGCGGCTGAGGCTGAGTGTGCTGCTGCTGCTGCTGCTG 6247
Oy 1621 CTGGAATGCTGCTGCTTTCGCGCTCTTCTGACCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 1680
Db 6248 CAGGTGAGCGCGGCTGCTCATGACCGTGGGAGATTCACCAAGGCTGATCAATCAG 6307
Oy 1681 CCGGCTGTTTTCGAAATGCGCGCTGCGGCAAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTG 1740
Db 6308 CCGGCTGTTTTCGAAATGCGCGCTGCGGCAAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTG 6367
Oy 1741 TCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1770
Db 6368 TCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6397

```

RESULT 5
 ABO16866/c
 LOCUS ABO16866 310029 bp DNA linear BCT 09-OCT-2003
 DEFINITION Pseudomonas syringae pv. tomato str. DC3000 section 11 of 21 of the

```

ACCESSION complete genome:
VERSION ABO16866
KEYWORDS ABO16866.1 GI:28853240
SOURCE Pseudomonas syringae pv. tomato str. DC3000
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
          Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 310029)
           Bell, C.R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I.T.,
           Gwinn, M.L., Dodson, R.J., Deboy, R.T., Durkin, A.S., Kolonay, J.F.,
           Madupu, R., Daugherty, S., Brinkac, L., Beanan, M.J., Haft, D.H.,
           Nelson, W.C., DavidSEN, T., Zafar, N., Zhou, L., Liu, J., Yuan, Q.,
           Khouri, H., Fedorova, N., Tran, B., Russell, T.V., Berry, K.,
           Uterbacher, T., Van Aken, S.E., Feldblyum, T.V., D'Acenzo, M.,
           Deng, W.L., Ramos, A.R., Alfaro, J.R., Cartimoun, S., Chatterjee, A.K.,
           Delaney, T.P., Lazarowitz, S.G., Martin, G.B., Schneider, D.J.,
           Tang, X., Bender, C.L., White, O., Fraser, C.M., and Collier, A.
           The complete genome sequence of the Arabidopsis and tomato pathogen
           Pseudomonas syringae pv. tomato DC3000
           Proc. Natl. Acad. Sci. U.S.A. 100 (18), 10181-10186 (2003)
           12928499
TITLE JOURNAL
PUBMED 2 (bases 1 to 310029)
REFERENCE Bell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,
AUTHORS Berry, K., Uterbacher, T., Van Aken, S., Feldblyum, T., Gwinn, M.,
           Dodson, R., Deboy, R., Durkin, A., Kolonay, J., Madupu, R.,
           Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J.,
           Nelson, W., DavidSEN, T., White, O., Fraser, C., and Collier, A.
           Direct Submission
           Submitted (03-MAR-2003) The Institute for Genomic Research, 9712
           Medical Center Dr, Rockville, MD 20850, USA
           Location/Qualifiers
FEATURES
source 1..310029
           /organism="Pseudomonas syringae pv. tomato str. DC3000"
           /mol_type="genomic DNA"
           /strain="DC3000"
           /db_xref="taxon:223283"
           /note="pathovar: tomato"
           204..1664
           /locus_tag="PSP02809"
           /locus_tag="PSP02809"
           204..1664
           /locus_tag="PSP02809"
           /note="similar to GP:15140142; identified by sequence
           similarity; putative"
           /codon_start=1
           /product="conserved hypothetical protein"
           /protein_id="AO56309.1"
           /db_xref="GI:28853241"
           /transl_table=11
           /product="peptide ABC transporter, permease protein"
           /protein_id="AO56310.1"
           /db_xref="GI:28853242"
           /translation="MIGWALRRRLTOSLIVLMTVVVFGVAINAGNPMIDLVGSDNLO
           ABLQIADIGLDPOIMCOYIFIKAGMAGNIGOSFVYEDARLILQIPAPPELAF
           SALFLAVVIGVPLGNFRAGMTPDHPLSRLLMAGSIVSPSPAPVALMILNLSITLGG
           LPAISGRGETREWLGVQSWLITDGLQHLILPALNLALFKISLVLRITRAGAVRVLDOE
           FVKEPARAKGISPMRVCMCHVMRNTMIPIVTVLAMELGSTIAVAVTESIFAMPAGKLI
           IDSINMLDRPVVAVIYLVVIFVILNLIIVDLVYILDRPVRIRASR"
           1661..2635
           /locus_tag="PSP02810"
           /locus_tag="PSP02810"
           1661..2635
           /locus_tag="PSP02810"
           /codon_start=1
           /transl_table=11
           /product="peptide ABC transporter, permease protein"
           /protein_id="AO56310.1"
           /db_xref="GI:28853242"
           /translation="MIGWALRRRLTOSLIVLMTVVVFGVAINAGNPMIDLVGSDNLO
           ABLQIADIGLDPOIMCOYIFIKAGMAGNIGOSFVYEDARLILQIPAPPELAF
           SALFLAVVIGVPLGNFRAGMTPDHPLSRLLMAGSIVSPSPAPVALMILNLSITLGG
           LPAISGRGETREWLGVQSWLITDGLQHLILPALNLALFKISLVLRITRAGAVRVLDOE
           FVKEPARAKGISPMRVCMCHVMRNTMIPIVTVLAMELGSTIAVAVTESIFAMPAGKLI
           IDSINMLDRPVVAVIYLVVIFVILNLIIVDLVYILDRPVRIRASR"

```

gene 2632. .3552
/locus_tag="PSPRO2811"
CDS 2632. .3552
/locus_tag="PSPRO2811"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AA056311.1"
/db_xref="GI:28853243"
/translation="MNVIKAVPANTLLAQSPRRIVDFMSSSTAVVGLLLLVITV
LLAAPMVIVONPYDLMQNLNMDAMPKSGSLNMSGYVWIGTGGGSDIYASITYG
LRISLWVIGSALIAAVITGLGVSAYAGWVDALIMLVLLSPYITLMAIMTLA
WKGKGVNMLTLLLEMAVYARTAGQALTESRKYVDAAQGVGPURIVIGTLP
NCLPLVIGALQIARATLEATLSFLGSGVPTSESLDLLIANGFQWMLSNBYWISL
FGLALITIVININLWGRRLRDVLPNRLQR"
3549. .4562
/locus_tag="PSPRO2812"
CDS 3549. .4562
/locus_tag="PSPRO2812"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, ATP-binding protein"
/protein_id="AA056312.1"
/db_xref="GI:28853244"
/translation="MSLPVSVSVTSVEPTLEVRDICTSFYTRAGVLPVAVNSLRVQ
PRLIIGLVGSGSKSVTPSILGLVDEGRISGGVLFKGRDLTKLSTGQUREIOGN
RAAMIFQDPMNTLPVLRVDTQMIKAVRAHSTGKREAREHARSRLALMGISPEERL
RAYPHOLSGMRORVAIALALHAPDLITADEPTTALDVITQQLISEYOKVIREGRT
SLIWTHTLSVAGLADVDVAVNAGRIIVGSGVADVLDROHRYTOGLIDSLPSRRKR
GGRLRQIPGMARDLSMPAGCAPARACAAAPACADPKPREILPGHTYRCYPCAGA
DVQ"
4552. .5565
/locus_tag="PSPRO2813"
/note="This region contains an authentic point mutation,
causing a premature stop, and is not the result of a
sequencing artifact; peptide ABC transporter, ATP-binding
protein, authentic point mutation; similar to GP:15158466;
identified by sequence similarity; putative"
5618. .7183
/locus_tag="PSPRO2814"
CDS 5618. .7183
/locus_tag="PSPRO2814"
/locus_tag="PSPRO2814"
/note="similar to GP:15140494, and SP:P23847; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, periplasmic
peptide-binding protein"
/protein_id="AA056313.1"
/db_xref="GI:28853245"
/translation="MKPLVRSLLASALVLAAGSVSAQODLRIGVADPVSSLDLPQNLNNA
GGRSVLHAFESLVSRRDDKTLPGAKLSKWTVDPTWEFALREDDVMQGTPLTADDL
VSPFERARVPGSVASVAGAMRTVESVKAKDHTLLIKRLNPANLLIPVDVSIYISR
HAGCAASSADVNSGKALIGTGPRYRPSFPGQRTIFARNDSSWGAPETMDKIDRRITA
NNAKRTALLAGDVLDIVKVSPTVBERLKTESVNFAYQALNLIQSFSTGPEER
IDNAGKPLAENPLLDVRRKALSLAINPALIDERMOGTVEALNPMANFPGVNG
INIPYDVQAKDLIAQAGFPEGFQLTVVPGRYEQAEVWQAVQAFYRIGVKKOL
EVLPMVAVGAKNKELIASVLANGTGEAAVALNLTIVDSSSGQASMSGHNSN
PLVDRKALADSTAEKDEAKRRKILDESVGVVDVGIILPFIHQNIMDAKGLKVEPLV
SPRTAAITWTEQP"
7255. .7776
/locus_tag="PSPRO2815"
CDS 7255. .7776
/locus_tag="PSPRO2815"
/note="similar to GP:15162228; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA056314.1"
/db_xref="GI:28853246"
/translation="MTTEHSNTPDDLHLGLIFPGNPMTYVHARPKVASATOGSAL
FDPALADNLPISERLWVAIVATQLSAQTLLATYTLQQLREVGVADSVLADVANRDI

gene 1126. .8366
/locus_tag="PSPRO2816"
CDS 1126. .8366
/locus_tag="PSPRO2816"
/note="similar to GP:15162229; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="alkyl hydroperoxidase Ahpd domain protein"
/protein_id="AA056315.1"
/db_xref="GI:28853247"
/translation="MSELIEIGFTNQVLGWRAMPPIVDNNAETAEQTALESHPQA
RTSDYVLTAAHHPDLRQSQAFNAIMVAPGGISAEELASTVSRINRCYCASH
AQRPOLAKRNDVIAOVFADPATGTAERKIVQFADITLEPASINAAHIIALREQ
GLDDAQVLDLJHAIAIPMANRLMLNIGSPVPEAE"
8429. .8614
/locus_tag="PSPRO2817"
CDS 8429. .8614
/locus_tag="PSPRO2817"
/note="similar to GP:15158870, and GP:15980012; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AA056316.1"
/db_xref="GI:28853248"
/translation="MAECHPSLFVNCARVSATYTFEPGARTAMTHPLGQTLIVTAG
ACQVDAEYASFPAPQ"
complement (8591. .9982)
/locus_tag="PSPRO2818"
CDS complement (8591. .9982)
/locus_tag="PSPRO2818"
/note="identified by match to PFAM protein family HMW
PF01569"
/codon_start=1
/transl_table=11
/product="fatty acid desaturase"
/protein_id="AA056317.1"
/db_xref="GI:28853249"
/translation="MYILVYGLLYFLVYHPVKSVRIVQPSGVDAHVPLPPTLPYL
STLVNPLVIVYKRGSSWILPTFFGALAAAGCLVCHLEFPMILRPETGAAMLDLY
QUDAPLAPSPSGHVALPAIVSAMAGLRLOKMAVWMSVIMLVMTVTGQHVFDWA
YGAAGVACGAATLIFKRCAYDLRTLSLLEMLGILTVRVALYLDWRFYLLAMLI
VVAROHALFVLVHDATHVNLTRRSSTNDFLIMALIGVPLVPIEFYRLPHLDHQAIG
TEODERRFLYRQPMRPOPLSAKLLRLQDLGDLNLRNIAVYKAGAPKPTR
PLISALVIMTIVACLWOCSTREVGMALMFIPLVYGTLLQKRSIAHSGGPGV
TFGQEWYTYMRVGNLGRFTIPYHINHLQHRTASLPMHALPSAVGKDEKLMPSRA
LPALMWSGLRRTY"
10166. .10954
/locus_tag="PSPRO2819"
gene
Query Match 54.5%; Score 965.2; DB 1; Length 310029;
Best Local Similarity 71.6%; Pred. No. 6e-221;
Matches 1267; Conservative 0; Mismatches 503; Indels 0; Gaps 0;
QY 1 ATGCGTTCTCTCGAGCGCTTCTCCCTCCCGGGTGAATTCCTGAGCGTCTCGAGCAT 60
DB 172246 ATGCGAGCCCTGGAACCGTTGGCGCCGTCGATCGGTCTGACACGGCTGGTGCAC 172187
QY 61 TGGGCTAAGACCCGTCAGAACAAACCTGCTGCTGCGAGGCGGCAATGGGGAGATGG 120
DB 172186 TGGGCGAGCGCTGTCGCAACAAACCTTATTTGCGCGCCGTGACAGCGTGGTGGCTGG 172127
QY 121 CGTCGATATGCTAGCGGGAATGTTCACCAACGTCGCGGCATGCGACAGAGCTTGCTT 180
DB 172126 CGCAAGGTACGTTAGCGCCGACATGCTCACCGAGTTCGGCATTTGCCAGAGACCTTGCTG 172067
QY 181 CCTTACGACATATCGGACAGAGCGTCCGCTGCTTATCGTCTCTGGAATATGACCTGGAACAT 240
DB 172066 GCTTACGAGCTTTCCGACAGAGCGGCCCATCTGGCGCTGCTGTCCGGTAAACGATCGAGCAT 172007

QY 241 CTTGAGCTGATTTGGGCTATGATGCGGCAATTCCTATTTGGCCGGTCTCTGCT 300
 DB 172006 TTGCAATTCCTTTGGGCTGATGATGCGGCTATTCCTATTTGGCCGGTCTCTGCT 171947
 QY 301 TATTCATGCTGCTGCAAGATTTGGCGAGCTGCTGCAATCTGAGTCTTTGCAACG 360
 DB 171946 TACTCGGTTCTGTCGAGGATTCGCGCAATTTGGGATGCTGATGAGTGTTCAGCA 171887
 QY 361 GAGCTGCTTTTGTGCTGCTGCAATGCAACCTTTTCCAGGCGCAATTTGACCAATTTCCG 420
 DB 171886 GGGCTGCTGTTTGTGCAAGCTGCTGCGCGCTTCCAGGCGCGCAATGCTGATCATCCG 171827
 QY 421 GAGCAGCTGCGCCCAATCTTCACTGCGAGGGAATTTGGCGGGGCGCGCAAGTGTG 480
 DB 171826 GCTGACACACCGGTCATTTACGATACGCGACAGTTGAATGCGCGCTCATTTGAGCTTT 171767
 QY 481 GACAGCTGCTGAGCAGCCTGCTGAGATTTGAGGCAATATATGCTTTGGCGCAACTGCG 540
 DB 171766 GCGAGCTGCTTGAACAAACCGGGGCGGAGGAGCGACGCTGCTCATTTGCGGAGCGG 171707
 QY 541 CCGGATACGATTTGCAAGTCTTTGATCTTTGCTGCTTCAACCAATGCTTAAAGCGGTG 600
 DB 171706 CCGGACAGCAATTTGCAAGTCTTTGATCTTTGCTGCTTCAACCAATGCTTAAAGCAAGT 171647
 QY 601 CCGACTACTGAGGAATGCTCTGCGCCCAATCAGAGATGCTTTCGCAACTTTCCGGTT 660
 DB 171646 GTTACACTGAGGCAATGCTGCTGCTGCTTCAACCAATGCTGCTGCAACTTTCCGGTT 171587
 QY 661 TTTGCTGAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
 DB 171586 TTTGGGGAACAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 171527
 QY 721 AGCCACACATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
 DB 171526 AGTCATATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 171467
 QY 781 CCAACCGCCCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 DB 171466 CCGACTGTTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 171407
 QY 841 TACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
 DB 171406 TACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 171347
 QY 901 CTGCGGGAAGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
 DB 171346 TTGCTGAGAGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 171287
 QY 961 CAGGGAATCGGATGCTTGAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 DB 171286 CAAAGTATCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 171227
 QY 1021 ATGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 DB 171226 ATGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 171167
 QY 1081 CTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
 DB 171166 CTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 171107
 QY 1141 GTGATGAGGAATTTGAAGGCGCTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 DB 171106 GTGATGAGGAATTTGAAGGCGCTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 171047
 QY 1201 GCTCTGGAACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 DB 171046 TCGCGGAGCAACCGCTGAGGCTTTCAGAGAAAGCTTTTATCTGTTCCGAGGAGCG 170987
 QY 1261 ATCAATTTGCAAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
 DB 170986 ATCAAGCTGCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170927
 QY 1321 GAGAGCTTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380

DB 170926 GAGATTTCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170867
 QY 1381 GTTCTGAGAGCGGCTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
 DB 170866 GTGCTGAGAGCGGCTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170807
 QY 1441 CTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
 DB 170806 CTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170747
 QY 1501 GAGGCTGCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
 DB 170746 CAGCGAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170687
 QY 1561 CTCAACGACTCAATCGAGAGCACTGCGCAATGCTGCTGCTGCTGCTGCTGCTGCTG 1620
 DB 170686 CTGCGACGTTTGAATGAAGGCGCAATGCGCAAGCTGCTGCTGCTGCTGCTGCTG 170627
 QY 1621 CTGATACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
 DB 170626 CTGATACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170567
 QY 1681 CCGCTGCTTTCGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
 DB 170566 CTGCTGCTTTCGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170507
 QY 1741 TCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1770
 DB 170506 TCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170477

RESULT 6
 LOCUS AE016786/c 301214 bp DNA linear BCT 12-DEC-2002
 DEFINITION Pseudomonas putida KT2440 section 13 of 21 of the complete genome.
 ACCESSION AE016786 AE015451
 VERSION AE016786.1 GI:26557030

KEYWORDS
 SOURCE Pseudomonas putida KT2440
 ORGANISM Pseudomonas putida KT2440
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE
 1 (bases 1 to 301214)
 Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,W., White,O., Peterson,J., Khouli,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzaz,A., Utlarback,T., Rizzo,M., Lee,K., Kosack,D., Moesti,D., Medler,H., Lauber,J., Honneisel,J., Straetz,M., Helm,S., Kiewitz,C., Eissen,J., Timmis,K., Duesterhoft,A., Tummeler,B. and Frazer,C.
 Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440
 Environ. Microbiol. 4 (12), 799-808 (2002)

TITLE
 JOURNAL Pseudomonas putida KT2440
 AUTHORS
 REFERENCE
 2 (bases 1 to 301214)
 Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,W., White,O., Peterson,J., Khouli,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzaz,A., Utlarback,T., Rizzo,M., Lee,K., Kosack,D., Moesti,D., Medler,H., Lauber,J., Honneisel,J., Straetz,M., Helm,S., Kiewitz,C., Eissen,J., Timmis,K., Duesterhoft,A., Tummeler,B. and Frazer,C.
 Direct Submission
 Submitted (05-NOV-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1..301214
 /organism="Pseudomonas putida KT2440"
 /mol_type="genomic DNA"
 /strain="KT2440"
 /db_xref="taxon:160488"

FEATURES
 source
 1..301214
 /organism="Pseudomonas putida KT2440"
 /mol_type="genomic DNA"
 /strain="KT2440"
 /db_xref="taxon:160488"

```
gene
CDS
    complement(69..1007)
    /locus_tag="PP3196"
    /product="complement(69..1007)"
    /locus_tag="PP3196"
    /note="similar to GB:X79440, SP:Q16798, and PID:535012;
    identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="modulation protein d1"
    /protein_id="AAN68804.1"
    /db_xref="GI:24984828"
    /translation="MRFHLDLNLVALDILIEEQNITPAERLHKTSQATSGVGLR
    LRFEBELLVQGRKMQPTPYALELAKPREVELLITRASITAPVPFDPAASKRHFV
    ASDYLISVFAQVIOKIQHQAQPIETPEMIGPDSALMRGEVMMIVPERYLIEGH
    PAOLLEEDHVCVVMRDNPEVGEHLTLBOYMOGHSVSGRNRHLSDIMFMNNOYGF
    NRRLVITNDPNTLPOLLVGTQVATMRRLAELVAGYLPRLPPVAKIPVWRERML
    MHRMDGDPMHMRKRISDVICASTQPDALRAT"
    1284..1225
    /locus_tag="PP3197"
    1284..1225
    /locus_tag="PP3197"
    /note="similar to GB:X54232, SP:P35052, and PID:31847;
    identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="glyoxalase family protein"
    /protein_id="AAN68805.1"
    /db_xref="GI:24984829"
    /translation="MNILGDLNLFVGVNDIQAACDLRDYGLDANOLDGGGFEALD
    GTSIIRRGDDDELPPALGPTTSRLRTYGVASSTDLAIADBLGDRVRRDASGL
    HFDFDGAFAIQVTCRPFDAADLTNAPGAPORINQLGIAPMPALPTLSHV
    YFVPDPAKGEAFYDRDLGFTVDSFIQVGFPKTAMDDHCLFMIQTPEPHQGCHEF
    TFMGSGTIVLLAGTRFGQCGWTSFMGPRHLGSNMFYFNSPLCGHIEYADMDKH
    DDAAMARQAPMADNSQLFLPASREKMSGPPAKQG"
    2238..2576
    /locus_tag="PP3198"
    2238..2576
    /locus_tag="PP3198"
    /note="identified by match to Pfam protein family HMM
    PF00355"
    /codon_start=1
    /transl_table=11
    /product="ferredoxin, 2Fe-2S"
    /protein_id="AAN68806.1"
    /db_xref="GI:24984830"
    /translation="MELCRDLDELSEGOARGFDPQGRGADSLFALRYLGQVRYRNLCP
    HLMVPLEYKRNRFSLNDGQWICYAHGARFTBEDGICVHGPCRGALDELDRKAGGW
    LLPLLEQDR"
    2716..3840
    /locus_tag="PP3199"
    2716..3840
    /locus_tag="PP3199"
    /note="similar to SP:P23262; identified by sequence
    similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="monooxygenase, putative"
    /protein_id="AAN68807.1"
    /db_xref="GI:24984831"
    /translation="MNTVNNVLIVGGGIGSLCAALIRRKGIANDVLEIKTEWTVYG
    GIIQGSNVREMRGLLEAYIDAAFAFEDVALITTEGOQLARIPEQRLAGPQYPMNV
    GISRLAHVLSSETAIELTOVRLGTSVESFEDDSADVLFIDSGQGRGLMVGADG
    LBSKYSVLFCDGYQPRFTGQSVRYNFPRAAGIDILANYQSGNAGLPLADIMLY
    LFTTSHPCNPMPNEPALADGRLORLAGFGLIGELRBOIVDNAGVYKLEAVFVDE
    PMYCGVILLIGDAAHATPHLIGGAGMALIEDAVVLAELTAGGSLDAQIQRFRGRFE
    RKFTSEKSLAGVKEIQRPSPFDRIGLVKRMLEVTAAQPI"
    3885..5597
    /locus_tag="PP3200"
    3885..5597
    /locus_tag="PP3200"
    /note="similar to GP:3184126, and GP:3184126; identified
    by sequence similarity; putative"
    /codon_start=1
```

```
gene
CDS
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AAN68808.1"
    /db_xref="GI:24984832"
    /translation="MPTPHNRKKEETPMWLRKGAFAAPRLPAKLPALVALAASQA
    QAEPDGTGNPDLRVRMDNLTKTSAMARTDASSKLTBQGTALNLDGDGDFPKGLISN
    RUDLSELDITTYNNVAGVSAWADVYNOGTDNNDPGRANSYSDYDAFTDDTRKL
    HGRKGEHDAFIFGKTEIGDMPLSGRLGQYAMQWGESLFGNNGLAGMAPYDVVYGL
    SVNPQFKEILIPVOOISGLOLQTPDVSVGAYQFEMANRLPAAGSYSTDDFEGDG
    ARMFVGAVPFPGGQPLAFPHGNDKOAKDSGGGQMVRRTERTEVDGLTAIRPHDKSP
    QIANRPNRANLIDNVSGKAGEFYWYVDEGEARGASPTTVGPNLAGELSTRWNOPLA
    STNQHALAAGEAIDDDNPVATGRTLHANVMSLSEPNFLAABESSFGEIAMNRVM
    SVSKADADVPADADATSLRVFEPMKQIIPGLDSVPGASVYNGASMLAGTGF
    TTHGGLNLTGLNGYNLTMSLDITTYTHYVGPENTFLADNHYTFEQSLKDRDIFATV
    STRP"
    7034..8095
    /locus_tag="PP3201"
    7034..8095
    /locus_tag="PP3201"
    /note="identified by match to Pfam protein family HMM
    PF02012"
    /codon_start=1
    /transl_table=11
    /product="BNR domain protein"
    /protein_id="AAN68809.1"
    /db_xref="GI:24984833"
    /translation="MNKMLYGMWLLAACGLPPAFAGDSYDVLAQPALHGPOLRAVLQ
    DVARADARLVAWGERGVLLSDNCAATWQGAFAVPGVLTAVQFIDARNGVAVGAG
    VTLHSDODGGEHVSVDLGDRAALAVOAAEAADRRLAAROLADGADKPLALNF
    TPARNGLVYGAIGVAMRTDDGSRMSQSEIGRLPNRGHLVYALAQGNVYIAGEBGL
    LTRSDGSHYHALGPPDGYSTFAATVLDGLJGLGSGSFPASDDGDSFOALANF
    TPASINGLRAGGQOLLNQAQMLRSLGSLGAAPAVASDGLPLTAATTAADGSIVA
    VMAARARLPESPLSTTTD"
    8108..10588
    /locus_tag="PP3202"
    8108..10588
    /locus_tag="PP3202"
    /note="similar to GP:9949187; identified by sequence
    similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="transporter, putative"
    /protein_id="AAN68810.1"
    /db_xref="GI:24984834"
    /translation="MNSAPSTGPIDDPDSAGSLERALLFNHRYVLILCLATLIL
    GNOARLTLNASEFKMIPRDHPFINHYLHROBELGLNAVLAVANPESGYXDKOYL
    HSLQOLNDAVYLLPVDRAAMKSLMTPSPRMTGVEEGEGPVDPVDGAAASIEA
    LKRNVERNEIQOLVAFDQRSSIVYVPLLEKTPDQALDYTPAHLETILRRRPOAG
    VEIHITGFKVVDLMDGLRQTLFFPAALAITAVALTWYTRCVRSITLVVCSLAV
    IWQGLPLGLYALDPYSVLPPLVFAIGMSGAGAKMGIMODIGMRHLVAARFTF
    RRLPLAGLTALLCDAVGFALMIIQIOVODLAVIASLGVAVLIFNLILPLVLSYV
    GVSARBARSLRAEBEASGAKHVMRFLDFTERRRMAACIAVALMAAGSYAVSL
    HLKVDLIDAGABELFADSRVYNDNAFYTRHVGASSDVPVAMRTAPGCGSAVGTLKH
    DDLIDQKRLGQVSDTNSLALNRYVLVGLSGSKPYDVLVNNQATLNMVTKAPRGL
    INDDCSLTLVAVLTDHKADTLARVVDSDQAANDSDQASFLMAAGSAGIBANQV
    VKQANRMDLMMWVGAVIVLCVTFRSMAVLCVAVLPVLTSLCEALMVAIGIGVQ
    ATLPIVATGCGIVGDIAYVMSIVLAQOGASLSQAYVRALLFTGKVMLTGTLTAI
    GVGTVIFSPIKFOADMGVLLAFMFWNMWVAILILPALAYFLLPRKACTEQLALPAH
    EHTRRERSGMEQASFRPLREVEHGR"
    10578..11768
    /locus_tag="PP3203"
    10578..11768
    /locus_tag="PP3203"
    /note="identified by match to Pfam protein family HMM
    PF04171"
    /codon_start=1
    /transl_table=11
    /product="major facilitator family transporter"
    /protein_id="AAN68811.1"
    /db_xref="GI:24984835"
    /translation="WVAERTAGSAQALILFGSCLPVGAIVLAPVLPMAHVAFTP
    GVAIVLPVALTIPALVIALIADLAGTADRVRRALILASMLISYSGILPLMVDLIV
    LVAVRAGIGLAEAGIMTCCTTLMGDIYFDGGRRARIPALOMVYTSILAAALFPGVGAAL
```

```

      GESDMPPTLVAVGALCLPLMLLLMEPRACOAASBASVDSRFPMAALPLVLLTVL
      AGVSLFIVPQAGYVLOLHIDA PROVGLAMGANOLGVALGALFUALPILPAGRLIA
      MGRATNGGGLMALASHPVPLVILINSLGVLPLITITVMOQVGDGGRATG
      GFTSALFAGEFVSPLVLALTAGVATPLPALILVALIGOLLAPVCLSLMRHRAVTV
      AGAR"
      gene
      11926..12513
      /locus_tag="PP3204"
      /locus_tag="12513
      11926..12513
      /locus_tag="PP3204"
      /note="Identified by Glimmer2; putative"
      /codon_start=1
      /transl_table=11
      /product="hypothetical protein"
      /protein_id="AA068812.1"
      /db_xref="GI:24984837"

Query Match      50.7%; Score 898.2; DB 1; Length 301214;
Best Local Similarity 69.4%; Pred. No. 7.6e-205;
Matches 1221; Conservative 0; Mismatches 538; Indels 0; Gaps 0;

QY      1  ATGCGTTCTCTCGAGCGCTTCTTCCCTTCCCGGTTGAATTCTTGAAGCTTCGAGCAT 60
DB      170844  ATGCAACCTGTCTGAGCCACTGGCGCGCTGCGCGCGCTGCTCGAGCGCTGTGTCAT 170785
QY      61  TGGGCTAAGACCGGTCCAGAACAACTGCGTTGCTGCGAGGGGCGCAATGGGGAATGG 120
DB      170784  TGGGCCACAGGTGGCGCGGACACCACTTTATGCGGACGCCAGGACGAGGCTGTGG 170725
QY      121  CGTCGTATCAGCTACCGGAAATGTTCCACAAGTCCGCGCATCGCAGAGCTTTCCTT 180
DB      170724  CGTTCATCAGCTACGAGAGATGCTGCGGATGTGCGACCAATCGCGCAACTTCTTA 170665
QY      181  CCTTACGAGCATACGCGAGAGCGTCCGCTCTTATCGTCTTGAAATGACCTGAAACAT 240
DB      170664  GGACTGGGCTTCACTGTCGCGCGCGCGCTGCGCTGCTTTCGCGCAACGACATGACAC 170605
QY      241  CTTGACGTGCAATTTGGGGCTATGTATGCGGGCATTCCTATTGCGCGGCTGCTGCT 300
DB      170604  CTGCAATGCGCTTCCGCGCAATGTATGCGGATTCCTATTGCGCGGCTGCTGCTGCT 170545
QY      301  TATTCACTGCTGTGCGAAGATTTGGCGAAGCTGCTGATCATCTGAGTCTTTCGCAACG 360
DB      170544  TACGCGCTGTGTGCGAAGACTTTCGCAAGTTGGCATGTGCGAGGTGCTCACCCCC 170485
QY      361  GGAATGCTCTTCTGCGGATGACGACACTTTTCAAGGCGCAATTGAGACATTTTCGCG 420
DB      170484  GGAATGCTCTTCTGCGGATGACGACACTTTTCAAGGCGCAATTGAGGCTGCTGAC 170425
QY      421  GACGAGTGCCTGCAATCTTCACTCGAGCGGAATTTGGCGGGGCGGACGAGTGTTC 480
DB      170424  GATTGCTGCGCGGTGATCAAGCTGCGGCGGAGTCCAGAGTGCCTCCCATATTAAGCTTC 170365
QY      481  GACAGCTGTGAGGACCTGTGTGAGATTTGAGGACATATATGCTTTTGGGCAACTGGC 540
DB      170364  GACAGCTGTGAGGACCTGTGTGAGATTTGAGGACATATATGCTTTTGGGCAACTGGC 170305
QY      541  CCGGATACGATTTCCCAATTTCTTGTTCATTTTGTGCTTTCACAACTGCTTAAGCGGTG 600
DB      170304  CCGGATACGATTTCCCAATTTCTTGTTCATTTTGTGCTTTCACAACTGCTTAAGCGGTG 170245
QY      601  CCGACTACTGAGGAATGCTCTGCGCAATCAGAGATGCTTTCGCAAACTTTCCGCGGT 660
DB      170244  ATACACACCGAGGACATGCTGTGCGCAATCAGAGATGCTTTCGCAAACTTTTCGAGC 170185
QY      661  TTTGTGAAAGCGCGCGGTGCTGTGAGACTGTGTCGCGTGAACCACTTCGCGGCG 720
DB      170184  TTTGTGAAAGCGCGCGGTGCTGTGAGACTGTGTCGCGTGAACCACTTCGCGGCGT 170125
QY      721  AGCCACAACATCGGCTGCTGTGTGTACAAAGCGGACGATCTTACCTTGAAGAGGTAA 780
DB      170124  AGCCACAACATCGGCTGCTGTGTGTGTACAAAGCGGACGATCTTACCTTGAAGAGGTAA 170065
QY      781  CCAACGCGCCAAAGGTTGCGCGAAGCGCTTTCGCAATTGAGCGAAATCTCTCCACATGCG 840

```

```

DB      170064  CCGACCCCGCAAGGCTTCCGCGAGACTTCGCGCATATGCGCGAGATTTCCCGACGGCC 170005
QY      841  TACCTCACTGTGCCGAAAGGCTGGAGAAATTAGTGGGTCCCTTGAAGCAGACATACC 900
DB      170004  TACCTCACTGTGCCGAAAGGCTGGAGAACTGTCAAGGACACTGAGCGAGACCCCGCG 169945
QY      901  CTGCGCAAGCGCTTCTTCCGTCATGAGACTGTTCTTCCGCGGCGGTGGGTGTGCG 960
DB      169944  CTAGCGAGGTGTTCTTTCGCGCATGAGCTGTTCTTTCGCGGCGGAGGCTGTGCG 169885
QY      961  CAAAGGATCTGGATGCTTTGACCGGAGTGTGTAACAGCATGCTGTGAGCGCATTCGC 1020
DB      169884  CAAAGGATCTGGATGCTTTGACCGGAGTGTGTAACAGCATGCTGTGAGCGCATTCGC 169825
QY      1021  ATGATGCGGCTGTGGGATGACGAGACTGCTCTTCTGCACTTTTACACCGGACCG 1080
DB      169824  ATGATGCGGCTGTGGGATGACGAGACTGCTCTTCTGCACTTTTACACCGGACCG 169765
QY      1081  CTGTCATGCTGCTTATCATTTGGGCTGCGAGCGCTGGGCTGGAGGTCAGCTGTTCGG 1140
DB      169764  TTGTCATGCTGCTTATCATTTGGGCTGCGAGCGCTGGGCTGGAGGTCAGCTGTTCGG 169705
QY      1141  GTGATGCGAAATTTGAAAGGCGCTTTCATGATTCGCGACGTCATGACGCGCTACTGCGT 1200
DB      169704  GTGATGCGAAATTTGAAAGGCGCTTTCATGATTCGCGACGTCATGACGCGCTACTGCGC 169645
QY      1201  GCTTCGTAACAATAATGCCCAAGCTTTCGACGAGAAAGCTATTACTGCTCCGATGATCC 1260
DB      169644  TCCCGCGACAGACCCCGCGAGGCTTTCGACGAGAAAGGCTATTACTGCTCCGATGATCC 169585
QY      1261  ATCAAAATGGCAATCTTCCGATTCCTCAGAAAGCTGATGTTTGAAGGTCGAAATGCT 1320
DB      169584  ATCAAAATGGCAATCTTCCGATTCCTCAGAAAGCTGATGTTTGAAGGTCGAAATGCT 169525
QY      1321  GAAAGCTTCAAGCTGCTTCGAGGAGTATTGTGACGCTTGGGCAATTGCGACGCGGCG 1380
DB      169524  GAAAGCTTCAAGCTGCTTCGAGGAGTATTGTGACGCTTGGGCAATTGCGACGCGGCG 169465
QY      1381  GTTCTGGAAGCGGCTTCTTACGTCCTGAGACGATGCTGCTCTGATTCGTAATGC 1440
DB      169464  GTTCTGGAAGCGGCTTCTTACGTCCTGAGACGATGCTGCTCTGATTCGTAATGC 169405
QY      1441  CTTGGAATGCTGCTTCCGCGCTTTCGACATGCGCGCTTGTGCGGAGTACGAAATA 1500
DB      169404  CTTGGAATGCTGCTTCCGCGCTTTCGACATGCGCGCTTGTGCGGAGTACGAAATA 169345
QY      1501  GAGGCGTCGAGCGCGAGGCTTTCGACATGAGCGCGGCTTGTGCGGAGTACGAAATA 1560
DB      169344  GAGGCGTCGAGCGCGAGGCTTTCGACATGAGCGCGGCTTGTGCGGAGTACGAAATA 169285
QY      1561  CTCAAACGATCAATGAGAACCACTGCGATATGCGATGCAATGAGGAGGCTC 1620
DB      169284  CTCAAACGATCAATGAGAACCACTGCGATATGCGATGCAATGAGGAGGCTC 169225
QY      1621  CTGATACGCGCGGCTGATGATTAAGGAGGAGTCACTGACAAAGGCTGATCAACAG 1680
DB      169224  CTGATACGCGCGGCTGATGATTAAGGAGGAGTCACTGACAAAGGCTGATCAACAG 169165
QY      1681  CGGCTGTTTGGCAATGCGGCTGCGGAAAGTATGATGCTGTATCTGTGTAAGATCAA 1740
DB      169164  CGGCTGTTTGGCAATGCGGCTGCGGAAAGTATGATGCTGTATCTGTGTAAGATCAA 169105
QY      1741  TCCATGCTGCTGACGAGG 1759
DB      169104  TCCATGCTGCTGACGAGG 169086

RESULT 7
PEU536324 4407 bp DNA linear BCT 09-JAN-2004
LOCUS PEU536324
DEFINITION Pseudomonas putida ech gene, vdh gene, fce gene and aat gene
(partial).

```

ACCESSION AJ536324.1 GI:40788111
VERSION at gene; beta-ketiolase; ech gene; fcs gene; feruloyl-CoA
KEYWORDS hydratase; feruloyl-CoA synthetase; vanillin dehydrogenase; vdh
gene.
SOURCE Pseudomonas putida
ORGANISM Pseudomonas putida
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1
AUTHORS Ficca A.G., Di Gioia D., Barghini P., Fava F. and Ruzzi M.
TITLE Identification of Pseudomonas putida strain T2 genes involved in
ferulate catabolism
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4407)
AUTHORS Ruzzi M.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2003) Ruzzi M., Agrobiology and Agrochemistry,
DABC - University of Tuscia, via C. de Lellis, snc - Viterbo,
01100, ITALY
FEATURES
source
1..4407
/organism="Pseudomonas putida"
/mol_type="genomic DNA"
/strain="T2"
/isolation_source="soil"
/db_xref="taxon:303"
/country="Italy:Lazio"
10..741
/gene="ech"
10..741
/gene="ech"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="feruloyl-CoA hydratase"
/protein_id="CAD60261.1"
/db_xref="GI:40788112"
/translation="MSPTLNREMDVLETVEDADARVVLITGEGSAMTAGMDLKEYF
REVDAPEILORIRRDASEMOWKILRFYSKPTIAVNGMCFGSGPLVACDLAICA
DEATGELSIINGIIPGNLVSXAMADTVGHRSLYIMGKTPDGOKAOMGIVNKS
PLAOLRDEVVLIAQDLIDKNPVLRPAKNGFKRCRLTMEQNEVDYIYALDQAEIDPE
HRAQGLKQPLDKSIKPELOAYKR"
782..2239
/gene="vdh"
782..2239
/gene="vdh"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="vanillin dehydrogenase"
/protein_id="CAD60262.1"
/db_xref="GI:40788113"
/translation="MPDVGALLIGKARASNDGFEFRBHPVSGEVSVAVAATLAD
ADAAVBAQAAPFPAALPGERBARKLAADLDRBEPFLAMASGTAKAMGFN
VMAANILREASMTTQIGSVIPSDVPSFPAALQPGGVVGLAPMNAPIVLAIRA
IAMPVACGTVVLKASEAPVHGLIGVYLAAGLGDGVVYVYCNAPHDAPAVERTLI
AMPAYRVNFTSGTHVGRIVGOLAARLKPVALEIGKAPFLVLAADLEAAVDAIA
FGAYVNOGQICSTERTLIVDQVADEFVAKIGEKVRLAADSOPDAPLGLSIDGA
GERIALVNDAYGKATIVAGERTGSVWQPVLLHYNAGMLYRESGSPAVYVIRA
SGDEBALALANDSERGLSAITFSRDTORLALAAQRLESGICHINGFTVADEOMPRGG
VVASGSGFSGAIDAFIQLRWVTVQHARHPYI"
2319..4175
/gene="fcs"
2319..4175
/gene="fcs"
/gene="fcs"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="feruloyl-CoA synthetase"
/protein_id="CAD60263.1"
/db_xref="GI:40788114"
/translation="MDCEFLSVAGRAPLRAVKIQGPQVQVROQGDVLYIDPVEITLDAY

PARLMEILLHMAQORSEQTFVARBAADGQWHTTAAMIQVRRIATWLLQGLSAR
PLILISGNDLEHVALATALVAGIPIYCPSPSPVSLISDQFGKLRHLYQLQELPVAA
DAAYGAIADVEFSPESINLIVGQGLAGMSTPFAEVLTVAPQCAQOAPALBPSLA
KELPFGSGTKLRAVTTTQGMLCANOMLLOJEPREAEVPLVDMLPNNHTFGGSH
VGLIVNGSTLYIDQKPTPOGPAOTHLQKIDISPAVLYIPKGMELVKALEODAL
CRRFTFORMLFFPAGAGLSQATWDRLDVAVAHACGERIRIMAGLCTERTSPSTPTTG
PLSVAGYIGLPAPEGCEVKLVPVGGKLEAFRGHVMVGRWSQOTQEPADDEGFCS
GDVAVRCDPNDPOLGIMPDGRJAEIDFKLSGVFVSVGPRQTVSSGSPYLVAVVIGP
DECEGLIIFPRLPECRQLSGAQOSTPKVLSPAVRCLSLATLARGATGNASY
IRFAGLAAPAPRIDRGETIDKSLNQRVVLQLRHGLIDSLVAGPRLVADAP"
4172..4407
/gene="aat"
4172..4407
/gene="aat"
/codon_start=1
/transl_table=11
/product="putative beta-ketiolase"
/protein_id="CAD60264.1"
/db_xref="GI:40788115"
/translation="MTARQFRTFATALLBGRTPWVDINGAFGLVSPDTDIGKAGR
AVLQRRRLDAGREVDVLAGSMAQADFADALLBRH"
ORIGIN
Query Match 43.1%; Score 763.2; DB 1; Length 4407;
Best Local Similarity 66.0%; Pred. No. 2e-172;
Matches 1136; Conservative 0; Mismatches 578; Indels 6; Gaps 2;
QY 11 TCGAGGCGCTTCTCCCTTCCGSGGTGGAATTTAGAGGTCTCGACATTTGGGCTAAGA 70
DB 2431 TGGAAACGTCGATGCTTACCCGACGCGCTATGAGCGCTTGCTGCATTTGGGCCAGC 2490
QY 71 CCGGTCAGAAACAACTGCTGCTGCGAGGCGGCAATGGGAATGGCGTGTATCA 130
DB 2491 AGGCAGCAGCAGACAGATTTGTGCGCGCGCGCTCCAGCGGCGATGGCAGACATCA 2550
QY 131 GCACCGGGAATGTTCCACACGTCCGCGCATGCGACAGAGCTTGCTTCAAGGAC 190
DB 2551 CTTACCGCCGCAATGTTGACAGCGTTCAGAGATTCGCAAGTGGCAGTGGCCAAAGGCC 2610
QY 191 TATCGGCAAGGCGTCCGCTGCTTATGCTCTGGAATGACCTGGAACATCTTCAAGCTGG 250
DB 2611 TGAAGCCGGAACCGCCGCTACTGATCTCTCGGCGACACACCTTGAGCAGTGCACCTGG 2670
QY 251 CATTTGGGAGCTATGATGCGGCGATTCCTATTTCCCGGTCTCTCTTATTCACTGC 310
DB 2671 CCACGCGCGCGCTGTATGCGGATTCCTATTTCCCGGTATCGCAGGATATTCACTGT 2730
QY 311 TGTGCGAATTTGGGAGCTGCGTCATCATGTAGTCTTCTGCAACCGGAGCTGTCT 370
DB 2731 TGTCCAGATTTTGGCAAGCTGCGCATATGCTCCAGACACTGCAACCGGAGCTGTCT 2790
QY 371 TTGCTCCGATGAGACCTTTTCCAGCGGCGCAATTGAGACATTTGCGGAGCAGAGTGC 430
DB 2791 TCGCGGCTATATGCG---CCTAGCGAGGCGCATCATGCGGTTTGCAACTCGATTA 2847
QY 431 CCGCAATCTTCACTGAGCGGAATTTGCGGCGGCGCGCACGCTGAGTTTGAAGCTGC 490
DB 2848 ACCGTGTGCTGGGCGCAGGCGAGCTGCGGCTTGGCGGAGCAGCGCATTTGCGGAGTAC 2907
QY 491 TGGAGACGCTGTGGGATTGAGCGAGATATAGCTTTGCGGCACTGCCCCCGATACGA 550
DB 2908 TGAACAACGTCGCCAACCTGTGCGGACAGAGCTTTGCGCGCGCGGACAGGCC 2967
QY 551 TTGCAAGTTCTTGTTCACCTTGTGCTTACCAAACTGCTAAGCGGTGCGCATACTC 610
DB 2968 TGGCAAGTTCTGTTTCACTTCCGCTCCAGCATGCGCCAGAGGCGGTGTGATCACACCC 3027
QY 611 AGCGAATGCTTGGCGCAATGACAGAGCTTCTTGAAACTTCCCGGTTTGTGTAAG 670
DB 3028 AGGCGATGCTGTGTGCAACAGCAGATCTGTGTCAGACTTCCCGGATTTGCGGAAG 3087
QY 671 AGCCGCGGTGCTGTGTGACTGTTGCGGTGGAACACAGCTTCCGCGGCGGACAGCAACA 730

Db	3088	TGCCGCGGCTGTGGTGCATCTGGCTGCGCTGGGAACCAACAGTTGGCGGACGCCAACAG	3147
QY	731	TCGGCATCTGTATTGTATCAACGGCGGCACAGTACTACCTTGAGACGGTAAACCAACCGCC	790
Db	3148	TCGGATCTGTCTGTATCAACACGGCGGCACGCTGTACATGACGATGTGGCAAGCCGACGCCG	3207
QY	791	AAGGTTTCGCGGAGACGCTTGGCAACTTGACGGAATCTCTGCCATCTGGATACCTCACTG	850
Db	3208	AGGCTTTTGGCCAGACGCTTGCAACAACCTCAAGAACATTTTCCCCACGGCGTACCTGACCG	3267
QY	851	TGCCGAAAGGCTGGGAGAGAAATTAGTGGGTGCGCTTGACGAGACAGTACCCCTGCGCAAC	910
Db	3268	TGCCCAAAAGCTGGGAGGAACTGTGTCAAGCACTGGACACAGAACCGCGCCCTGTGCGAGC	3327
QY	911	GCTTCTCTGCTGCGATGAAAGCTGTTCTTCTTCGCGGCGGCTGGTGTGTGCAAGGAATCT	970
Db	3328	GGTTCTTTCACAGCGCATGAAGACGTGTTCTTTTTCGCGGCTGCGGGTGTGTCCAGGCGCATCT	3387
QY	971	GGGATCTGTTTGGACCCGGTTCGCTGAACAGCACTGTGTGAACGCAATTGGCATGATGGCGG	10310
Db	3388	GGGACCGGCTTGAAGCCGCTGGCCGTGGGCCCACTGGGCGGAACGATTCGCAATGATGGCGG	3447
QY	1031	GTCGTGGGATGACGGAAGACTGCTCTCTTCGTCGACTTTTACACACGGGACCGCTGTGATGG	10909
Db	3448	GGCTGTGGGATATACCGAAGACTTTCGCGCTGTGTGCACTTTTACACCGGGCGCGCTGTGCTGG	3507
QY	1091	CTGGTTACATTGGGCTGCCAGCCGCTGGCTGCGAGGTCAAGCTGCTTCCGTTGCAATGGGA	11506
Db	3508	CCGGCTACATCGGCGCTGCGCGGCCCGCGGGGTGGAGTCAAGCTGGTGTGCGGTGACGGCA	3567
QY	1151	AATTGGAAAGGCGTTTCCATGTCGCGACGTCATAGAGCGGCTACTGTGGCTGCTCTGAAC	12110
Db	3568	AGCTTCAGGCGCAGAGTTTTCGCGGCGCCCACTGTATGTCGCGGGTACTGGCGCTCATTCGACG	36227
QY	1211	AAAAATGCCCAAGCGTTTGCACGAGGAGGACTTATCTGCTCCGCGGTATGCCATCAATCAATTGG	12707
Db	3628	AGACCCAGGAGGCGTTTGCACGAGAGGAGGCTTCTACTGACGCGTGTATCGGTGCGCTGT	3687
QY	1271	CAGATCTCTGCGCATCTCTCAGAAAGGTCTGATGTATTGACGCTGCAATTGCTGAGACATTCA	13310
Db	3688	GCGACCCCGGCAACCCGACGCTGGGTTTGAATGTTGACGCGGCGCATTCGCCAGACTTCA	3747
QY	1331	AGCTGTCTCAAGGGTATTTGTCAAGCTTTGGGCCATTGGCGCACGCGGCGGTTCTGGAA	1397
Db	3748	AGTTGTCTCTCGGGGTGTTGTGTCAAGCTTTGGGCCGATGCGCACCC--AGGTATCTTACG	3804
QY	1391	GCGGCTCTTAAGTCTCTGGAACGTAGAGGTGTCGTCTCTGATCTGTAATAGCTTGGATTCG	14506
Db	3805	GGTGCGCTTAATGTGTGTGATGTGTGTGTGTGGGCGCGGACCGCAATAGCTTGGGTTTTCG	3866
QY	1451	TCGTGTATTCCGCGTCTTCTGCACTGCGCTGCTTCTGCGGGCTAGGAAAGAGCGTTCGG	15110
Db	3865	TTATCTTCTCCCGCTGTGCGGAAATGCCCCGAGTTGTCCGGGCTTGGCCCCACAGGCACTA	3922
QY	1511	ACGCGGAGTCTTTCGCACTGAGAGCCGTTTCGGGCTGTGTTGTGCACTGGCTCAAAACGAC	15707
Db	3925	CGCCCAAGGTCTGGCATCTGCCCCGCGTACGCTGTCGCTGGCCAGCTGTCGACAGC	3984
QY	1571	TCAATTCGAAAGCACTGGCAATGCGAATCGATATGATGGGTAGGGCTCTTCATATACG	16310
Db	3985	TCAACCCGGGGGCGCGACAGGCAACGCAACTCATCTCCGCTTGGCCGCGCTGGCCGCGAAG	4044
QY	1631	CGCGGTGATTTGATTAAGGCGAGGTCTCATGACAAAGGCTGCATCAACAGCGCGCTGTTT	16910
Db	4045	CACGCGCCATGACCGGTGTGATGATCATCGACAAAGGCTTCCTCAACACGCGGTGTCC	4104
QY	1691	TGCAATGGCGGTTCGGCGAAAGTTGATGGCTGTGTATCGTGG	17310
Db	4105	TGCAATTCGGGACGCGCTGATTCGACAGCTGTATGACAG	4144

LOCUS	189050 bp	DNA	linear	BCT 11-APR-2003
DEFINITION	Ralsstonia solanacearum GM11000 megaplasmid, complete sequence;			
ACCESSION	AL646077			
VERSION	AL646077.1			
KEYWORDS	GI:17430642			
SOURCE	Ralsstonia solanacearum			
ORGANISM	Ralsstonia solanacearum			
AUTHORS	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; Ralsstonia.			
TITLE	1 Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangelot, S., Ariat, M., Billault, A., Brotier, P., Camus, J.C., Catolico, L., Chandler, M., Chistone, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaepin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Sigulier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weissensbach, J. and Boucher, C.A.			
JOURNAL	Genome sequence of the plant pathogen Ralsstonia solanacearum			
MEDLINE	Nature 415 (6871), 497-502 (2002)			
PUBMED	21681879			
REFERENCE	11823852			
AUTHORS	2 (bases 1 to 189050)			
TITLE	Boucher, C.A.			
JOURNAL	Direct Submission			
COMMENT	Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA UMRV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex			
FEATURES	Christian.Boucher@toulouse.inra.fr http://seqsequence.toulouse.inra.fr/R.solanacearum.html.			
SOURCE	Location/Qualifiers			
repeat_region	1..189050			
gene	/organism="Ralsstonia solanacearum"			
CDS	/mol_type="genomic DNA"			
	/strain="GM11000"			
	/db_xref="taxon:305"			
	/plasmid="megaplasmid"			
	79..1500			
	/note="RS49+ or RS06146			
	LONG DIRECT REPEAT 49 +			
	predicted by Homology"			
	/evidence=not_experimental			
	83..2971			
	/gene="RSp0175"			
	/note="synonym: RS04687"			
	83..2971			
	/gene="RSp0175"			
	/function="miscellaneous; hypothetical/global homology"			
	/note="Product confidence: putative			
	Gene name confidence: hypothetical			
	predicted by Codon usage			
	predicted by Homology			
	predicted by Framed"			
	/codon_start=1			
	/evidence=not_experimental			
	/transl_table=11			
	/product="PUTATIVE VGR-RELATED PROTEIN"			
	/protein_id="CAD17326.1"			
	/db_xref="GI:17430643"			
	/db_xref="GOA:Q8XTD8"			
	/db_xref="TrEMBL:Q8XTD8"			
	/translation="MDKTKLQHLFAPAHRLVYALGEGP1AELAVEAMLGREALSELE			
	EMRWVSAANARFALISFCGRVTLVTTLDGTSARSTGLRQAEQAGADSLARYR			
	TVWVPLATQORHSGOVNRPPLADILEOVVSPAPYAAWMPFAGAEDRNAPFGRSR			
	IAORERTYRFTVRLIALAGIGFTVDEGAPSGHALLIFADSPRLAEDTSAAGGI			
	RYRHSQOEEDDAITLIGSRVAVGVAAWAPQAKRAFRVAVPSRFGCIASDPD			
	YLTSPSLAPDTANAQRIALQVMSVVARAPLFTGRGDSVRLRSGTRKTVDCPLP			

[illegible]

```

gene      .7527
          /gene="RSP0178"
          /note="synonym: RS04690"
          6745..7527
CDS
          /gene="RSP0178"
          /function="miscellaneous; unknown"
          /note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=11
          /product="HYPOTHEITICAL PROTEIN"
          /protein_id="CAD17329.1"
          /db_xref="GI:17430646"
          /db_xref="GOA:Q8XTD5"
          /db_xref="JTBMBL:Q8XTD5"
          /translation="MTQYLYHTITTAVARIIIRKGLTLAAHPALGRPVARRGAEV
NRAEVRGROVNRLKAVIKKGLAGSLDIRGQRPFI PVVPAGNRDDEVEIIR
VERAEVKAFIALATGAPANKRGRLTLPRTIGEHADMLRRKANALCRLAHVHSLEY
AIIEGMTSRHVSREPRASDCSYTRQGCAQCSYLVSFMAAAALPLDDSDRA
VTMRKLIPQOILEIMRAPSDVLTFTNADDRNAAGNMPLTQMS"
          7518..8021
gene
          /gene="RSP0179"
          /note="synonym: RS04691"
          7518..8021
          /gene="RSP0179"
          /function="miscellaneous; hypothetical/partial homology"
          /note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=11
          /product="HYPOTHEITICAL PROTEIN"
          /protein_id="CAD17330.1"
          /db_xref="GI:17430647"
          /db_xref="GOA:Q8XTD4"
          /db_xref="JTBMBL:Q8XTD4"
          /translation="MEILNAPPDMGOPTRMVAVRGAAILALSALLAAPALADARKPMQ
DEPARAAVSEARQYEYEVSGTHIPLYLIHADPAMVAVQSTVDVYHRNAVILPGSKSLI
GKELRKVARREIVWIGLOIPLAASGTURLDPLOATWRDSSAGVADISPEALIGAMTS
EPPIVPH"
gene
          complement(8066..8488)
          /gene="RSP0180"
          /note="synonym: RS04692"
          complement(8066..8488)
          /gene="RSP0180"
          /function="miscellaneous; hypothetical/global homology"
          /note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Codon_usage
Query Match      41.9%; Score 741.4; DB 1; Length 189050;
Best Local Similarity 64.2%; Pred. No. 3,6e-167;
Matches 1130; Conservative 0; Mismatches 626; Indels 3; Gaps 1;

        2   TGCCTTCACGAGCGCCCTTCCTCCCTCCGGGTGCAATTCTTGAGCGCTCGACCAT 61
            |||||               |||||               |||||               |||||
Db    66492  TGCGCTCGGTGAGCGCATCGGTCCTTCGCGCGGACAGCGCGCGCGCATATGGCGA 6655
        62   GGCGTTAAGACCCTGCCAGAACAACCTGCGTGTGTCGACGGCGCGCAATGGGGAATGC 121
            |||||               |||||               |||||               |||||
Db    66552  GGCGCGGACAGCATCCGCGATCGGTCCTTCGCGCGGACAGCGCGCGCGCATATGGCGA 6661
        122  GTGCTATCACACTCGCGGGAATGTCACAAACGTCGCGCGCATCGACAGAGCTTGCTTC 181
            |||||               |||||               |||||               |||||
Db    66612  AGCGCATCACCTATGCCCAATGCTGCACCGGCGCGCGCGGTGCTGCGCAGGCGCTGCTGG 6667
        182  CTTAACGACTATCGCGAGACCGTCCGCTGCTTATGCTCTTGAAAAAGACTCTGAACATC 241

```

Dh	66672	CGCGGGGGCTGTGCGCCGAGCGGCGCTTGCTGATTCCTGTGCGGCAACGATCTGCAACATC	66731
Qy	242	TTCAAGCTGGCATTTTGGGGCTATGTATCGGGGCAATTCCTCATTTGCCGGGTCTCTGCTT	301
Dh	66732	TGCAGCTGGCGCTGGGGCGCCATGTATATCGCGCAATTCCTCATTTGCCGGGTGTCGCGCCCT	66791
Qy	302	ATTCACTGCTGTGSCAAGATTTTGGCGAAGGTGGGTCAATCGTAAGTCTTTCTGCAACCGG	361
Dh	66792	ACGGCGTGGTATCGCAGAGACTTACAGCCGGCTGCTACTACTGTGTGCGCACTTGAACGCGG	66851
Qy	362	GACTGTCTTTGTGCTGCGGATGACAGCACTTTCCAGCGCGCAATTGAGACCAATCTTGCGG	421
Dh	66852	GCCCTGTCTATGCCACCGACGGTGTGCTGTTCCCGCGCGGATCTCAGCGGCTGTGTCCCG	66911
Qy	422	ACGAGTGGCCCGCAATCTTCACTCGAGCGCAATTGCGCGGCGGCGCCACGCTGAATTTTG	481
Dh	66912	CGGAGACCGAGAGTGCAGATGCACCAACGCGAGAGGTGGCGGGGCGCGCTGTACCCGGCTGG	66971
Qy	482	ACAGCCCTGTGGAGACACCTGTGTGGATTGAAGCAGATTAATGCTTTGGCGCAACTGGGC	541
Dh	66972	CGTGCCTGTGCG---CGACCGGGGCGGCGCAGCTGTGATGCGCAACGCGCGGTGGGCG	67032
Qy	542	CCGATAGCATTTGCCAAGTTCTTTGTTCACTCTGGCTCTACCAACTGCTTAAGCGGCTGC	601
Dh	67029	CCGACAGATCGCCCAAGTTCTCTGTACCTCTGGGCTCGACGCGCAGCCCAAGCGGCTCA	67088
Qy	602	CGACTACTCAGCGAATGCTCTGCGCCAAATCAGCAGATCTTCTGCAAACTTTCGCGTTC	661
Dh	67089	CCACACAGCATTCGAGTCTGTGCGAGCAACAGCAGATGCTGTGCGAGACTTTTCCCTGCT	67148
Qy	662	TTGTGTAAAGACCGCGCGGTGCTGTGTGATCTGTGCTGCTGCTGCTGCTGCTGCTGCTG	721
Dh	67149	TGCGCGAGGCGCGCGCGCTGTGTGATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	67208
Qy	722	GCCCAACATCTCGGATGCTGTGTGTATCAACGGCGGCGCACTATCACTTTCGCAACGCTAAC	781
Dh	67209	GCCCAACGTTGGGATGCTGTGTATCAACGGCGGCGCACTATCACTGACGAGCGCAACG	67268
Qy	782	CAACCGGCCCAAGGTTTCGCGAGACGCTTTCGCACTTGAACGCAATCTCTCCACTGCGT	841
Dh	67269	CGAACCGCAGGGCTTTCGACACAGCCTGTGCGCAACTGTGCGCACTGTGCGAGCGGATGT	67328
Qy	842	ACTCTACTGTGCGGAAGGCTGGGAGGATTTAGTGGGTGCTTTGAGCGAGACGATACC	901
Dh	67329	ACTTCAACGTGTCCCAAGGGCTGGGAGATGCTGACCGACGCGCTGACGCGCAACCCGCGCC	67388
Qy	902	TGCGCGAAGCTTTTGGCTCGCATGAAGTGTCTTTCTTCCGGGCGGCTGTGTCTGCG	961
Dh	67389	TGCGCGAGCACTTCTATGCGCGCTGCGGTGTCTTCTTCTTGTGCGGGCGCGGCTGTGCG	67448
Qy	962	AAGGGATCTGGGATCGTTTGGACACGGGTCGCTGTAACAGCACTGTGTGAGCGGATTCGCA	1021
Dh	67449	AGCGCGGCTGGGACCGGCTCGAGGGCGTGAACGAGGCCCATCTGGCGAGCGCATTCGCA	67508
Qy	1022	TGATGGCGGGTCTTGGGATGACGAGACTGCTCTCTTCTGCACTTTTACACGACCGCG	1081
Dh	67509	TCAATGCGCGGCTTGGGATGACGAGACCGCGCGCTGTGTCACAGTTTCAACACGCGCCCG	67568
Qy	1082	TGTGTATGCTGTGTTATCAATTTGGGCTGCGACGCGCTGTGCTGCGAGGTCGAAGCTGTTCGG	1141
Dh	67569	TCATGATGCGCGGCTATCATGCGCTGTGCGCGCGCGCGCTGCGAGGTCGAAGCTGTTCGG	67628
Qy	1142	TCGATGGGAAATTGGAAGGGCGTTTCATATGTCTCGACGTCATAGACGGGCTACTGCGCTG	1201
Dh	67629	TCGACGGGAGCTGGAAGCCCGCTTCAAGGACCGCATGTTCATGCGGGCTACTGCGCGG	67688
Qy	1202	CTCTGTAACAAAATGCCCAAGCGTTTCGACGAGAGAGGCTATTACTCTCGGTATGCCA	1261
Dh	67689	CGCGGAGCTGGCGCGGTGAAGCGTTTCGACGAGACGCGCTATCACTGCTGGGCGAGCGCG	67748
Qy	1262	TCAAAATTGGCAGATCTCTCGGATCTCTGAAAGGCTGTGATGTTTGAACGTTGCAATTGCTG	1321
Dh	67749	TGCCTTCTGTGACGAGCGCGCGCCGACAGTGTGCGCTGATGTTTGAACGCGCGCATGCGCG	67808

FEATURES	source	Location/Qualifiers
gene		1..346879 /organism="Rhodopseudomonas palustris CGA009" /mol_type="genomic DNA" /strain="CGA009" /db_xref="taxon:258594" 127..1005 /gene="dapa1" /locus_tag="RPA1571"
RESULT 9		
LOCUS	BX572598	346879 bp DNA linear BCT 20-APR-2004
DEFINITION	Rhodopseudomonas palustris CGA009 complete genome; segment 6/16.	
ACCESSION	BX572598 BX571963	
VERSION	BX572598.1 GI:39648490	
KEYWORDS	complete genomes.	
SOURCE	Rhodopseudomonas palustris CGA009	
ORGANISM	Rhodopseudomonas palustris CGA009 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Rhodopseudomonas.	
REFERENCE	1 (bases 1 to 346879)	
AUTHORS	Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S., Tabita, F.R., Gibson, J.L., Hanson, T.E., Bobst, C., Torres, J.L., Peres, C., Harrison, F.H., Gibson, J. and Harwood, C.S.	
TITLE	Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris Nat. Biotechnol. 22 (1), 55-61 (2004)	
JOURNAL	2 (bases 1 to 346879)	
PUBMED	Larimer, F.W. and Harwood, C.S.	
REFERENCE	Rhodopseudomonas genome consortium	
AUTHORS	Direct Submission	
CONSRTM	Submitted (24-JUL-2003) Submitted on behalf of the Rhodopseudomonas genome consortium, the DOE Joint Genome Institute, Production Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA, and the Genome Analysis Group, Oak Ridge National Laboratory, 1600 Commerce Park Drive, Oak Ridge, TN 37831, USA;	
JOURNAL	1600 Commerce Park Drive, Oak Ridge, TN 37831, USA; larimerf@ornl.gov	

CDS
127. .1005
/gene="dapA1"
/locus_tag="RPA1571"
/EC_number="4.2.1.52"
/function="InterPro IPR002220 COGS COG0329"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative dihydrodipicolinate synthetase"
/protein_id="CAE27012.1"
/db_xref="GI:39648491"
/translation="MMADFPVPPYIVSPVDAGRVRADVMGLCDLLOAGVHGLTP
LDSGFAYLIGTAQREAVRATIEAQRVPVAGVASTVADAVAQALYELGLDGI
IIAIEAYFPLKDAQIESYFRAIADAVEIPVYITNPQORSLTLDVIALEHPR
RIKADSTWGRLLSIIINRCGDAIOVSASAHIPAAVMILIGVMAGPACIAPROSV
ALYELCKAQRMDPAIMLQRLKMRVNEAPAFKFNLAACIKGMLAQGVDPPIPPQAL
TAEERKAVKVLAEIA"
1230. .3560
/locus_tag="RPA1572"
1230. .3560
/locus_tag="RPA1572"
/function="InterPro IPR000674 COGS COG1529"
/note="Observed by proteomics
Citation: Proteomics from Verberkmoes et al. (2003)
unpublished"
/evidence=not_experimental
/codon_start=1
/transl_table=11
/product="putative carbon-monoxide dehydrogenase large
subunit"
/protein_id="CAE27013.1"
/db_xref="GI:39648492"
/translation="NNILPGSMKFGAQGVKRLDQRLLTGHGGLYDDKAEGALMLV
VNRSPAHADVISIDTEAKGMGVPAVLTGADLIADVGITPLPIPRPDSPPAL
PIRLIILAHKVRPVEGPPVAAYVATSOIAQAAMEAIAVEYHEIPATDPPAALQAP
ATVETPNDIVNASTGYDAKVDAAASASHTSIDLVSGRLFSMERSTIAIEK
KVRLLHTQSQTPGQTRDALAEAILKRPRESIQVLVSGIGGFGQKTVYEPDAIVA
YAAVKLNRIKRMGERIDETIGTGRDLTSTFALDSKGLAVRVISIGGTGYL
AGAGVIIPVLGPFQVGVHLPVHFDIKAWTHTAPVGAAYGAGRPEAVYVERLM
DAAKRLMDPRAIRKINYIKPSQLPYNAVGVVGAFAHMDRAELSGDGLDKKA
RKKAQKGLVGRGVTSYIEWGSRAHNTYHATAGRGVILHSGTAMGGGLT
YTMIAALEIPLDKIDVQGNDDLAIGSVSSRLPVGTRVAVASAMTKAKAEK
AAHTEASIEDIEYSGMLTIAGTRKISLFTIAAEKGRLSVDSITGEVDGSPWNG
AHICEVVDPEGVSRVRYTVDVGNVPMPLVAGQIHGVAGVQALYENAAVN
DGQQLTASYLDYICIPRADNLPIINVLDPSPAPCKTNPLGAKGSGSGAIGBPVCH
GVDAIAPLGVTTLNTPITEKVKWRAIQAKSAQAA"
3638. .4252
/locus_tag="RPA1573"
3638. .4252
/locus_tag="RPA1573"
/function="COGS COG1704"
/note="observed by proteomics
Citation: Proteomics from Verberkmoes et al. (2003)
unpublished"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="LemA family"
/protein_id="CAE27014.1"
/db_xref="GI:39648493"
/translation="MRKFLTVLALALSLNSGNYTTIQSEDEVKSTSEVVNQYR
RADLVNVLNVSXGFAQOBERDVLGKTNARAKGSIQATPEVINDPAAPQOQGE
LSSALSRLLIVTENYPOLKSDLEFKMLMAQLBETENRITVARNRYIKAAVAAYTVRS
VPNTFAMMEGVYKKNFTYENKAIISTAPKVDFTAPAPAPAK"
3638. .3700
/locus_tag="RPA1573"
/note="Signal predicted by SignalP 2.0 HMM (Signal
peptide probability 0.998) with cleavage site probability
0.396 at residue 21"
4257. .5180
/locus_tag="RPA1574"
4257. .5180
/locus_tag="RPA1574"
/locus_tag="RPA1574"

misc_feature
sig_peptide
sig_peptide
misc_feature
gene
CDS
/function="COGS COG1512"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="DUP477"
/protein_id="CAE27015.1"
/db_xref="GI:39648494"
/translation="MTAAITRAVRAGMALLCVVALAPAPRADVAAPVLTGRVVD
QGTLSADAVLADQIKAFPEARKGSIIVLIVPTTPPAIEQFSIRVAEMKIGRKK
IDDGAILVAARDKRLRIEYVGLGALPDVAKRIIDETIPKFTGDPGGIEAGV
DMIGVIGDEPPTPQOHEMAAPESMDTGMATPDLFGTIVNGLFTRALGRVAS
GITGIVGAVNMLFGLVYIHALVAGRAFLILVIDLFGPTTSSGRRGSGGTWIG
SSGSSWSSSSSSGFSGGSGSGGASGSW"
4257. .4352
/locus_tag="RPA1574"
/note="Signal predicted by SignalP 2.0 HMM (Signal
peptide probability 1.000) with cleavage site probability
0.947 at residue 32"
0.947 at residue 32"
order(4290. .4358,4824. .4880,4899. .4952,4962. .5030)
/locus_tag="RPA1574"
/note="4 probable transmembrane helices predicted
by TMHMM2.0"
5194. .5691
/locus_tag="RPA1575"
5194. .5691
/locus_tag="RPA1575"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE27016.1"
/db_xref="GI:39648495"
/translation="MGIRIGKHLFTRGAVHRAFPDPAIERAIKASETHVQGI
RVVEGALGADLPFRQDIPVERALDIFSLRIWDTFNNNGVILYLALARNVYIIDR
GIDRGAGTWEAICREMAEFREGEFEGVLRGIDRIHAHLAEHPRSGSPNEIPD
APVVV"
complement(5773. .6384)
/locus_tag="RPA1576"
complement(5773. .6384)
/locus_tag="RPA1576"
/function="InterPro IPR004045 IPR004046 COGS COG0625"
/note="observed by proteomics
Citation: Proteomics from Verberkmoes et al. (2003)
unpublished"
/evidence=not_experimental
/codon_start=1
/transl_table=11
/product="putative glutathione S-transferase"
/protein_id="CAE27017.1"
/db_xref="GI:39648496"
/translation="MKLYDSIGPNPRLVMTFAEKGLITPTQLDIRAGENRKPEHIA
RPHGOTPLLELDNGSFLEAVTAICYLEBEKHOPPLIGTPEERAECPEMTIRIDLY
ICEPLNGRFAEGKMKQDRVVCVPEARGLKTIAADRIWMNEGMKGEYCGARL
TIADLLIYGLDLPGAOVQPLAPANTTIEMFKRIARPSAKA"
6593. .7564
/locus_tag="RPA1577"
6593. .7564
/locus_tag="RPA1577"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE27018.1"
/db_xref="GI:39648497"
/translation="MDEHNEPIRGEOGTAAATDGAONAGTPBASASAPOGTTPAN
TPANCAKFPDEAKPRAGKSEIKSDRSVPRGTYTTPASPRREDIYRMWDEIIEVA
PDBASQSAVHRGVRKLAVAVAIVIAIAGVAGLATASSHRSDBATTAASAVYK
TIDAOVQAOIEISALKDLIDHAQKOSTIGKTDRIKLEKASQSPAKAKISBT
VEKLRTAQVQPPAAPANVAAPKADVGSITPKKAAPKGLPTVEGVLADVDVG
GAVBGRSGTVEYVADPPIGVGRVDAIRQDGRMVVTVSRGLIVAR"
6947. .7015
/locus_tag="RPA1577"
/note="1 probable transmembrane helix predicted

```

terminator      7593..7621
                /note="TransTerm prediction, 98+%"
gene            complement(7721..8494)
                /gene="fpr"
                /locus_tag="RPA1578"
                complement(7721..8494)
                /gene="fpr"
                /locus_tag="RPA1578"
                /EC_number="1.18.1.2"
                /function="InterPro IPR001433:1PR001834 COG6 COG1018"
                /codon_start=1
                /evidence=not experimental
                /transl_table=11
                /product="Ferredoxin--NADP+ reductase"
                /protein_id="CAE27019.1"
                /db_xref="GI:39648498"
CDS
Query Match      29.8%; Score 527.8; DB 1; Length 346879;
Blast Local Similarity 56.6%; Pred. No. 7.7e-116;
Matches 976; Conservative 0; Mismatches 747; Indels 0; Gaps 0;

QY      2 TGGGTTCTCGAGGCGCTTCTTCCCTCCGCGGTCAATCTTGAAGCTTCGAGATT 61
DB      256342 TGGGCTGACACACACCGCTACCGACTATCCGGTGGGATTACGACCGGCTGATCACT 256401
QY      62 GGGCTAAGACCCGTCAGAACTGCGTTCGCGAGGCGGCAATGGGAAATGCG 121
DB      256402 TCCCGGAGACGGCGCCGACCGGGTGTCAATGCGCCGAGGAAAGGTGAGGCGCGCTGGC 256461
QY      122 GTGGTACAGCTACGCGGAAATGTTCCAAAGTCGCGGCAATGACAGAGCTTGGCTTC 181
DB      256462 GCAAGATACGATACGAGAGTGTGCGCGCGGCAAGACATGCGCTCGCGCTGATCG 256521
QY      182 CTTACGACTATCGGAGAGCGTCCGCTGCTTATGCTCTGGAATGACCTGGAAATC 241
DB      256522 CGGCGGAGCTGTGGCGGAAACGGCGGGTGAATGTTGTCGGTAATTCATGATCAATG 256581
QY      242 TTCAAGCTGGCAATTTGGGGCTATGATAGGGGCAATTCCTATGCGCGGTCTCTGCTT 301
DB      256582 CGATGGTGAATGTTGGGCGCGCTGTATGCGGGCGTGGAGTGTCCGGTGTGCGCGCGT 256641
QY      302 ATTCAGCTGTGCGAAGATTGGCGAAGTGGCTCAATGTAAGTCTTTCGACCGG 361
DB      256642 ATTGCTGTGTCTCAAGGATTTACGCAAGCTGCCCATATGTCGGGCTGTGACTCGG 256701
QY      362 GACTGTCTTGTGCTGCGATGACGACCTTTCAGCGGCAATTGAGACCAATTCGCGG 421
DB      256702 GGGTGAATCTTCGCGGATGACATCACTGCTTCGCGCGCGATCTCGGCAACGCTGGCG 256761
QY      422 ACGACGTGCCCGCAATCTTCACTCGAGCGAATTTGGCGGCGGCGGCAAGTGAATTTTG 481
DB      256762 AGGATGTGAGCTGCTGCCCGCGCGGCGAGTGAAGGGGCGCAAGGTGACGTGCTTG 256821
QY      482 ACGAGCTGTGAGACGCTGCTGGGATTTGAGCGAGATTAATGCTTTGGCGGCACTGGCC 541
DB      256822 CCGAATCTGTGCGCAACGCGGAAATCCGAACTCGCGCCAGACAGAAAGCATTCGCC 256881
QY      542 CCGATACGATTTGCAAGTTCTTGTCTTCTGCTCTACCAATGCTGCTAAGCGGCTGC 601
DB      256882 ACGACACATTCGCTAAGTTTGTGCTGAGTGGGATGACCGGCAATTCGAAGCGGCTGA 256941
QY      602 CGACTACTCAGCGAATGCTTGGCGGCAATCAGCAGATGCTTTCGCAAACTTTCCGGTTT 661
DB      256942 TCATATACGACGCGGATATCTGCCCAATCAGTGAATGATCCGAGAGCGATGCGCTTCC 257001
QY      662 TTGGTGAAGGCGCGGCTGTGTTGATGATGCGGTGGAAACACACCTTTCGCGGCA 721
DB      257002 TGAAGAAGACGCCCGGTGATGCTCACTGCTGCGGTGAATACACTTTCGCGGCA 257061
QY      722 GCCACCAATTCGCGCATGCTGTGTTGACAAAGCGGCGGCACTACTTACCTTGAACGAGTAAC 781
DB      257062 ACCACAAATTCGCTGACGCTGTTCAACGGCGGCTGATGTATATGACGACGCGCAAGC 257121
```

```

QY      782 CAACCGCCAAAGGTTGGCCGAGACGCTTGGCAACTTGAAGGAATGTTCCCACTGGCT 841
DB      257122 CGAGCGCGCGCGGATGCTTCCACATTCGCAATTCGCGGAATGCGCGACGGTGT 257181
QY      842 ACCTCACTGTGCGCAAGAGCTGGAGAGAAATTAATGAGTGGCTTTCAGAGACAGTACCC 901
DB      257182 ATTTCAACGTTCCGAAGGCTACGAGTGGCTGTGCGCGGTCTGCGGAAGACAGCAAGT 257241
QY      902 TGGCGAACCTTCTTCTGCTCGCATGAACTGTTCTTTCGCGGCGGCTGGGTTGTGCG 961
DB      257242 TGGCAATATGTTCTTCAAGCGGCTGATGATGTTCTTCTCGGCGCCAGCTCGCGG 257301
QY      962 AAGGATCTGGAGATCTTTTGAACGGGTCCCTGAACAGCACTGTGTGAAGGCAATTCGCA 1021
DB      257302 CGCATGTCTGAAAGGAGCTTTGACAGAGTCCGGGTGCGGAGAACCGGCGCGGGTGCGA 257361
QY      1022 TGATGCGGGTCTGGGATGACGAGACTGCTCTCTCTGCACTTTTACACCGGACCGC 1081
DB      257362 TGCTCACCGGCTTGGCGCCACGAGACCGGCGCTTCTTCAATGTCGATGACCGCGAGA 257421
QY      1082 TGTGATGCTGTGTTACATTTGGGCTGCGCAGCGCTGCTGCGAGTCAAGCTGTTCCG 1141
DB      257422 CCAATGCTCGGCGCATGTGCGCTGCGGATGCCGCAACGAGGCAAGCTGTGTCGGA 257481
QY      1142 TCGATGGAAATTTGGAAGGCGCTTTCATGATGTCGCAACATATAGGCGCTACCTGGCGTG 1201
DB      257482 ACAAAGGCAAGCTGGAAGTCCGCGCAAGAGGCGGAAATCAACCCCGCTATTGGGCGG 257541
QY      1202 CTCTGTAACAAAATGCGCAAGCGTTGACAGAGAAAGGCTAATTAAGTCTCCGGTATGCA 1261
DB      257542 CGCCGAGCTGACCGGATTAAGGCGTTTCAGACAGAGAGGCTTTCACACTCAACATGCGC 257601
QY      1262 TCAATTTGCAAGATCTGCGGATCTTCAGAAAGGTCTGATGTTTGAAGTGAATTTGCTG 1321
DB      257602 TGAAGCGGTGATGCAACCAAGACCTTTCGCGCGCTTCGATTCGAGCGCGGATTCGCG 257661
QY      1322 AAGACTTCAAGCTGCTCTCAAGGGGTATTGTAAGCGTTGGGCGCATTCGCAACGCGGCGG 1381
DB      257662 AAGACTTCAACCTGAGCTGCGGACCTGGGTGAGCGTGGGCTGAGCTACGCGCAAGTTCA 257721
QY      1382 TTCTGGAAGGCGGCTCTTAACTCTGAGAGTATGATGTTGCTGCTCTGATGTAATGCC 1441
DB      257722 TTGCGGCTGCGCTCTCCCTGCTGGTGGCGCATGATGATGCGCGGCTGACCGGATTAAG 257781
QY      1442 TTGATTTCTGTGTTTCCGCTTCTTCACTGCTGCTGCTTGTGCGGCTTGAAGAAAG 1501
DB      257782 TCACCGGCTGGCGATCTCGATCCGACGAGCTGCAAGCTGATCAATGCGACGCTGCGC 257841
QY      1502 AGGCGTGGACGCGGAGTGTCTTCCAGTGAAGCCGCTTCGCGGCTGTGCTGATGCTGGC 1561
DB      257842 TGGAAAGCTTCGCGGATGCGGCTGACCAATCTGATTCGAGAGGCTTCGCGAGGCGCT 257901
QY      1562 TCAACGACTCAATCGAAGCAACTGGCAATGCGCAATGCGCATGATGAGTGAAGGCTCC 1621
DB      257902 TCGCAGACGCTGTGACGCAAGGCAAGCGGCTGCTGCAACCGGCTACCCGCGCGCTGTGCG 257961
QY      1622 TCGATACGCGCGCTCGATTTGATTAAGGCGAGAGTCACTGACAAAGGCTTGAATCAACAGC 1681
DB      257962 TCGGCGAACCGGCTGTGATGACAAAGGAGATCAACGCAAGAGGCTCGGTCAACAGC 258021
QY      1682 GCGCTGTTTTCGAATGCGCGGTTCGCGGAAAGTTGATGCGGCTGTA 1724
DB      258022 GGGCGGTGTGAATATTCGCGCTGTGATTCGCGGATCTTTA 258064
```

RESULT 10
AP005963 299950 bp DNA linear BCT 12-SEP-2003
LOCUS Bradyrhizobium japonicum USDA 110 DNA, complete genome, section
DEFINITION 29/31.
ACCESSION AP005963 BA000040
VERSION AP005963.1 GI:27355938

KEYWORDS	REFERENCE
SOURCE	Bradyrhizobium japonicum USDA 110
ORGANISM	Bradyrhizobium japonicum USDA 110
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium.
AUTHORS	1 Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasaemoto, S., Watanabe, A., Idegawa, K., Iriyuchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpou, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110
JOURNAL	DNA Res. 9 (6), 189-197 (2002)
MEDLINE	22484898
PUBMED	12597275
REFERENCE	2
AUTHORS	Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasaemoto, S., Watanabe, A., Idegawa, K., Iriyuchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpou, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110 (supplement)
JOURNAL	DNA Res. 9 (6), 225-256 (2002)
MEDLINE	22485002
PUBMED	12597279
REFERENCE	3 (bases 1 to 299950)
AUTHORS	Kaneko, T.
TITLE	Direct Submission
JOURNAL	Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; 2-6-7 Kazusa-kamateri, Kisarazu Chiba 292-0812, Japan
MEDLINE	(E-mail:kaneko@kazusa.or.jp, rhizobase/, URL:http://www.kazusa.or.jp/rhizobase/, Tel:81-438-52-3935 (ex.2338), Fax:81-438-52-3934)
PUBMED	Location/Qualifiers
REFERENCE	1. 299950
AUTHORS	/organism="Bradyrhizobium japonicum USDA 110"
TITLE	/mol_type="genomic DNA"
JOURNAL	/strain="USDA110"
MEDLINE	/db_xref="taxon:224911"
PUBMED	36. .278
REFERENCE	/gene="bbsr7655"
AUTHORS	36. .278
TITLE	/note="ORF ID:bbsr7655
JOURNAL	hypothetical protein"
MEDLINE	/codon_start=1
PUBMED	/transl_table=1
REFERENCE	/protein_id="BACS2920.1"
AUTHORS	/db_xref="gi:27355939"
TITLE	/translation="MMVQATDVSDSSAIDENVAAVVDRFGRLDVVNNAGVHEGAPAS
JOURNAL	SITDKKRTVMSTVDGVFYSCRAALPTLKRPGRS"
MEDLINE	275. .637
PUBMED	/gene="bbsr7656"
REFERENCE	275. .637
AUTHORS	/gene="bbsr7656"
TITLE	/note="ORF ID:bbsr7656
JOURNAL	hypothetical protein"
MEDLINE	/codon_start=1
PUBMED	/transl_table=1
REFERENCE	/protein_id="BACS2921.1"
AUTHORS	/db_xref="gi:27355940"
TITLE	/translation="MNRVSVSGTDGDMGMSPPYNAKGVAVVNLTRALADLGKGIKRVNKA
JOURNAL	AVCSLRTGTEDMDKELLAKFAERIPLGRVCEPEVNAVAIAPLASDASPMTGAK
MEDLINE	NVAQSLGVASNSGCPQE"
PUBMED	complement(660. 1346)
REFERENCE	/gene="bbsr7657"
AUTHORS	complement(660. 1346)
TITLE	/gene="bbsr7657"
JOURNAL	/note="ORF ID:bbsr7657
MEDLINE	putative phosphoglycolate phosphatase"
PUBMED	/codon_start=1
REFERENCE	/transl_table=1

gene
protein_id="BACS2922.1"
/db_xref="cgi:27355944"
/translation="MMDWDESELPKAIIPDLDTGLDSDVLDHALMGEAVYERGHDT
FEVNSQIQGKKDKLIPVLSDEQOREHGRHLEWRNRNRFKSTYLPILRPSASQLL
QVRVRSQILRVASSAKEDELKKYLDIAAISDLVDVATSSRDVDSKPADIPFVYIK
KLGIQGSAAVAGIDTPYDAIARAKAGVAIGVLGCGFTENSLRAGCEVYPGPGTLY
ARFGPTATLK"

CDS
complement(1503. .1880)
/gene="b117658"
complement(1503. .1880)
/gene="b117658"
/note="ORF_ID:b117658"
/codon_start=1
/transl_table=11
/product="two-component response regulator"
/protein_id="BACS2923.1"
/db_xref="cgi:27355942"
/translation="MSNALRVLVVEDEALICSFIEDALSDGFPACSVSHGAAMSTP
RDGRQCALITDVNLGDIGISGWEIARQINRITGFPVYWTASAPDWKSGQVDGL
LIEKFPAAQIAAAVSQLDTR"

gene
complement(1987. .2319)
/gene="b117659"
complement(1987. .2319)
/gene="b117659"
/note="ORF_ID:b117659"
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BACS2924.1"
/db_xref="cgi:27355943"
/translation="MAMEDRPTSLISDNVEGNNVYAGGEKGYTERVMDIKVSK
VSHAVTSGDLIRIGDVHYPLPMQALKTDTNLGCVYTGITDHLRSAPKADSSNMN
SDPAFR"

gene
complement(2397. .2825)
/gene="b117660"
complement(2397. .2825)
/gene="b117660"
/note="ORF_ID:b117660"
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BACS2925.1"
/db_xref="cgi:27355944"
/translation="MKRRPRMAKFCRLIDTTIYVPVLRLOGAAARCLAGTLDGR
ACRQSAVAEIRSVIQPQTEBPQIPESIRGRSRSPVDFDLAGMRFGFHEHLQ
PGHDIMVYHRRAIDPGFAQLPQKFAQLQGGFPHARTPLQN"

gene
complement(2979. .3422)
/gene="b117661"
complement(2979. .3422)
/gene="b117661"
/note="ORF_ID:b117661"
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BACS2926.1"
/db_xref="cgi:27355945"
/translation="MLMQGEFFFAFNYPFLDPGMRVSLRPMRWDFPYDESGRPVSL
VRLRLIRORDHILRGCTYGFPEFNDMRVQFGILLFARHGAQYTLIATNFGDTRTV
PFWPKANAYVEELHVALDKNVPALQWLDIPSHYGRWVTLT"

gene
3427. .3654
/gene="b8r7662"
3427. .3654
/gene="b8r7662"
/note="ORF_ID:b8r7662"
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BACS2927.1"
/db_xref="cgi:27355946"
/translation="MPLERASGGSIAKLGSRWHLGRYYGGGRISVHGKEAVHYKANGG
SHGRFPFRAGPEAKRRSSSSSPGDIQRDM"

gene
complement(4162. .4824)
/gene="b117663"

```

CDS
    complement(4162..4824)
    /gene="b117663"
    /note="ORF: b117663
    unknown protein"
    /codon_start=1
    /transl_table=1
    /protein_id="BAC52928.1"
    /db_xref="GI:27355947"
    /translation="MKRMRYASISSTVIANGVCAALVVGQSHVPEPDIAASVTRAP
    ELMERAMHLPVAATFORHVDQMNGSRGCPAAVANAARSLGEAARTBGKVLNAGTSCM
    TGVCIMGLTIDELAKVAQNTSRKVTILRLDSENGPLEHLERANDPGRATYVNDPRAR
    IFGAGSGHSPIGGYFEAEDLVFLDVNFNPQWLVBRKRLFDAVNTLIDBKXKGLL
    IE"
    complement(7319..9061)
    /gene="b117664"
    complement(7319..9061)
    /note="ORF: b117664
    putative adenylate cyclase 3 (EC 4.6.1.1)"
    /codon_start=1
    /transl_table=1
    /protein_id="BAC52929.1"
    /db_xref="GI:27355948"
    /translation="MTEQORVQRRLAATIAADVGVSAIMQRAEATYAEFERLKRRLI
    EPSLSRHDGRILKTTGDGALAEFAPSLAAVCAEMODHLAEGSPKRLRIGLMIGV
    IYVODGLXGDGNIATVRLBEGIAAPGSLMESEKYSVEYEGLDYGFEDRGGQOLKITA
    KPVRAIVRAGASALTERLSAPLPKPSIAVLPENNSGDPEQSFADMGVEBIT
    TALSRFFMLFVIAINSSFTYKGAIVNKKQVGHGIVYILLEGSRKAGKRVRIAGOLI
    DAIQTGTHIMADPERDLIDIFALQDVALAVASVOPMQLQTEIATARREPMVLTAY
    DFLSRMQOYVLTRESLAELIKLHRALELDPPSSAALAGCHSNVALGYANNP
    EEPDAKEAIRLARLALSLDDGPDPLSRVAIVASACVGDSEAEIEMADRAVALNNSMR
    TYNRGVYTMAGLPBEAIRISFRBARIKMSPLDPLIHTFPLIGIHALVIGLRFDALITV
    LKALRLNPSFMPACRGLASFAHLGRDAEAAHAAVLEIDPTFTLSGHIAKRVKSA
    KMTLEGKRLKGLPE"
    9440..10444
    /gene="b117665"
    9440..10444
    /note="ORF: b117665
    putative methyltransferase (EC 2.1.1.-)"
    /codon_start=1
    /transl_table=1
    /protein_id="BAC52930.1"
    /db_xref="GI:27355949"
    /translation="MISPMWTLTGLNGFOVTOAIHVAATLRIADHLKDSRAEEL
    APLTQSHPGALYRLRLAALAVGFEDEDRFALTPEMDCIARTSATPIGMAACVSS
    PYWSTWGHLLHSIRTEGNAFOSLNGKDIWNYRAERPEESAVPRAMTEFRGAEAV
    IGAVDFAFHVVVDGGRGLMLAAILTAHPRMGGTLFDQGVLAGAABVALEAGVVD
    RCMVSGSPFESVPEGDAYVMRVYIHDMDDEAIALIKACRAMRETRAKLVLERI1

Query Match      27.7%; Score 489.6; DB 1; Length 299950;
Best Local Similarity 55.2%; Pred. No. 1.2e-106;
Matches 954; Conservative 0; Mismatches 774; Indels 0; Gaps 0;

gene
CDS
    complement(4162..4824)
    /gene="b117663"
    /note="ORF: b117663
    unknown protein"
    /codon_start=1
    /transl_table=1
    /protein_id="BAC52928.1"
    /db_xref="GI:27355947"
    /translation="MKRMRYASISSTVIANGVCAALVVGQSHVPEPDIAASVTRAP
    ELMERAMHLPVAATFORHVDQMNGSRGCPAAVANAARSLGEAARTBGKVLNAGTSCM
    TGVCIMGLTIDELAKVAQNTSRKVTILRLDSENGPLEHLERANDPGRATYVNDPRAR
    IFGAGSGHSPIGGYFEAEDLVFLDVNFNPQWLVBRKRLFDAVNTLIDBKXKGLL
    IE"
    complement(7319..9061)
    /gene="b117664"
    complement(7319..9061)
    /note="ORF: b117664
    putative adenylate cyclase 3 (EC 4.6.1.1)"
    /codon_start=1
    /transl_table=1
    /protein_id="BAC52929.1"
    /db_xref="GI:27355948"
    /translation="MTEQORVQRRLAATIAADVGVSAIMQRAEATYAEFERLKRRLI
    EPSLSRHDGRILKTTGDGALAEFAPSLAAVCAEMODHLAEGSPKRLRIGLMIGV
    IYVODGLXGDGNIATVRLBEGIAAPGSLMESEKYSVEYEGLDYGFEDRGGQOLKITA
    KPVRAIVRAGASALTERLSAPLPKPSIAVLPENNSGDPEQSFADMGVEBIT
    TALSRFFMLFVIAINSSFTYKGAIVNKKQVGHGIVYILLEGSRKAGKRVRIAGOLI
    DAIQTGTHIMADPERDLIDIFALQDVALAVASVOPMQLQTEIATARREPMVLTAY
    DFLSRMQOYVLTRESLAELIKLHRALELDPPSSAALAGCHSNVALGYANNP
    EEPDAKEAIRLARLALSLDDGPDPLSRVAIVASACVGDSEAEIEMADRAVALNNSMR
    TYNRGVYTMAGLPBEAIRISFRBARIKMSPLDPLIHTFPLIGIHALVIGLRFDALITV
    LKALRLNPSFMPACRGLASFAHLGRDAEAAHAAVLEIDPTFTLSGHIAKRVKSA
    KMTLEGKRLKGLPE"
    9440..10444
    /gene="b117665"
    9440..10444
    /note="ORF: b117665
    putative methyltransferase (EC 2.1.1.-)"
    /codon_start=1
    /transl_table=1
    /protein_id="BAC52930.1"
    /db_xref="GI:27355949"
    /translation="MISPMWTLTGLNGFOVTOAIHVAATLRIADHLKDSRAEEL
    APLTQSHPGALYRLRLAALAVGFEDEDRFALTPEMDCIARTSATPIGMAACVSS
    PYWSTWGHLLHSIRTEGNAFOSLNGKDIWNYRAERPEESAVPRAMTEFRGAEAV
    IGAVDFAFHVVVDGGRGLMLAAILTAHPRMGGTLFDQGVLAGAABVALEAGVVD
    RCMVSGSPFESVPEGDAYVMRVYIHDMDDEAIALIKACRAMRETRAKLVLERI1
```

```

Qy 302 ATTCACTGCTGTCCAGAGATTGGCGAAGCTGCTCATGTGAGTCTTTCGACCCG 361
Db 199581 ATTCGCTGTGTCCAGAGATTGGCGAAGCTCTCTATTCGATGAAGCTTTGACGCCG 199640
Qy 362 GACTGTCTTTGCTGCGGATGACACCTTTCCAGCCGCGCAATTGAGACCAATTGCGCGG 421
Db 199641 GCTGTCTTTGCTGCGGATGACACCTTTCCAGCCGCGCAATTGAGACCAATTGCGCGG 199700
Qy 422 ACGACGTGCCCCGAATCTTCACTCGAGCGAAATTGCGCGCGCGCGCACGGTGAATTTCG 481
Db 199701 TCGGACCGAGATGCGCGCTCTCAACGCGCATGTGGCGGCGGACGTCAGCTGCTCG 199760
Qy 482 ACAGCTCTGTGAGACACCTGTGTGAGATGAGACGATATGCTTTGCGGCACTGCGC 541
Db 199761 CCGACCTCAATGCGACCGCCGATCCGCGGATCTCGACGAAGTCCACGCAAGATCGCGC 199820
Qy 542 CCGATACGATGCGCAATGCTTGTTCACCTTCTGAGCTTACCAAACTGCTTAAGCGGTGC 601
Db 199821 CCGATACGATGCGCAATGCTTGTTCACCTTCTGAGCTTCAAGCTGAGATCCCAAGCGCTCA 199880
Qy 602 CGACTACTCAGCGAATGCTTGTGCGCAATCAGACGATGCTTTCGAAACTTTCCGCTTT 661
Db 199881 TCAACACCCAGCGCATGATCTGCGCAACAGGTGATGTGCGGAGACCTCGCTTCC 199940
Qy 662 TTGCTGAAAGACCCCGCTGTGTGATCTGTGTGCGGTGCGGAAACCAACCTTTCGCGCA 721
Db 199941 TCAAGGACGAGCCCGCTGTATCATCATGTGCTGCGCTTGAATCAACTTTGCGCGCA 200000
Qy 722 GCCACACATCGGATGCTGTGTGATGACAGCGCGGCACTACTACCTTGAACGAGTAAAC 781
Db 200001 ACCCAATATCGCTCAACGCTCTCAATATGCGCGCTGATGATATCTGAGCGCGGCAAGC 200060
Qy 782 CAACCGCCCAAGGCTTTCGCGAGACGCTTTCGCACTTGAAGCAATCTTCCACTGCGT 841
Db 200061 CGATGCCCGGATGATGAGAGACGCTGCGCAATCTCAAGAGATTTTCGCGACCTGCT 200120
Qy 842 ACCTCACTGTGCGCAAGGCTGTGAGAGAAATTAGTGGGTGCCCTTGACGAGACGATACC 901
Db 200121 ATTTCAACGTGCGCAAGGCTGTGAGAGAAATTAGTGGGTGCCCTTGACGAGACGAGGCC 200180
Qy 902 TGCGCGAAGCGCTTGTGCTGCTGCAATGAAGCTGTCTTTCGCGCGGCTGGTGTGCTGC 961
Db 200181 TGCGCGGAAATTTTTCGACCGGCTGCAATGCAATGCTTCTCGGCGCGCGCTGTGCTGC 200240
Qy 962 AAGGATCTGAGATGTTTGGACCGGCTGCTGAACAGCACTGTGTGAGCGCAATTGCGCA 1021
Db 200241 CGTTGCTGTGAGACGCTGCAAGAGCTCGCGGTGAAGAAAGGCTATGCGTGC 200300
Qy 1022 TGATGGGCGGTGTGAGGATGACGAGACGCTGCTTCTGCACTTTTACACCGGACCGC 1081
Db 200301 TGCTGACCGGCTGTGAGGATGACGAGACGCTGCTTCTGCACTTTTACATCGGATCGCGCA 200360
Qy 1082 TGTGATGCTGTGATGATGATGAGGCTGCTGAGCGCTGCTGAGAGGTCAAGCTCGTTCGG 1141
Db 200361 CCAAGCGCTCGGCAATGTGCGCTTCCCTGCTCGGCAACGAGCCAACTGATGCTCGCA 200420
Qy 1142 TCGATGGAAATTTGAAAGGCGTTTCATGATCGGCACTGATGAGCGGCTACTGCGTGC 1201
Db 200421 ACAACGCGCAAGCTGAGAGTCCGCGCAAGGCGCGCAAGTATGCTGCTGCTGCGCGC 200480
Qy 1202 CTCTGTAACAAATGCGCAAGCGCTTTCGAGAGAGAGGCTATTAAGCTCGGTGATGCA 1261
Db 200481 AAGCGATATCAACGCAAAATCTTTGAGAGAGAGGATTTCAAGATGAGGCGACCGC 200540
Qy 1262 TCAATTTGAGAGATCTGCGGATCTGAGAAAGGTGTGATGTTTGAAGCTGCAATTGCTG 1321
Db 200541 TCAAGCTGCGGATCTGCGGATCTGAGAAAGGTGTGATGTTTGAAGCTGCAATTGCTG 200600
Qy 1322 AAGACTTCAAGCTGTCTCAAGGAGTATTTGTCAAGCTGTGAGCGCAATGCGGCGG 1381
Db 200601 AAGACTTCAAGCTGTCTCAAGGAGTATTTGTCAAGCTGTGAGCGCGGCGGCTTTCG 200660
```

QY 1382 TTCTGGAAGCGGCTTTACGCTCTGAGACGTAGTGTGCTCTGATCGTGAATGCC 1441
Db 200661 TTCTGCGGCTCTGCGACCGCTGTGTGGCGACGTCTGTCATCGCCGGCATCAACCGGACGAGG 200720
QY 1442 TTGGATTTGCTGCTGCTTTCCCGGCTTCTCGACTGCGCTGCTTGTGGGGCTAGAAAAAG 1501
Db 200721 TCTCCGGCTGTGTGTGCTGTGACCTCGACTGTGTAGGCTGTCAACCCGACGCTGCGG 200780
QY 1502 AGCGCTGAGACGCGGAGGTCTTGCACGTAGCGCGGCTTGGGCTTGTGTGTGACTGAGC 1561
Db 200781 CCGACGACCTCTGCTGTGTGCGACCGGCGACCGGCTGTGTGCGGAGGCTTTCGGAGGCGC 200840
QY 1562 TCAAGCACTCAATCGAAGAACAACTGTGCAATGCGACTGCTATCATGTGTAGGAGCTCC 1621
Db 200841 TGACCCGCTTCTCTGCGCTCGGCGCACCGTTCCTGACGCGGATCACCGCGCATCTCTGA 200900
QY 1622 TCGATAGCGCGGCTGTGATGATTAAGGGCGAGGTCACTGACAAAGGCTTCATCAACGAC 1681
Db 200901 TGAACACGCGCTCTCTCATGACAAAGGCGAGGTCACTGACAAAGGCTTCATCAACGAC 200960
QY 1682 GCGCTGTTTTCATGAGCGGTGCGGCAAGTTGATGCTGTATCGTG 1729
Db 200961 GCGCGGTGTGTGAGCATCGGCGCGCTGTATCGACGAGCTCTACGCTG 201008

RESULT 11
AR319843 1902 bp DNA linear PAT 17-AUG-2003
LOCUS AR319843
DEFINITION Sequence 2393 from patent US 6562958.
ACCESSION AR319843
VERSION AR319843.1 GI:33700946
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1902)
Breton,G. and Bush,D.
Nucleic acid and amino acid sequences relating to Acinetobacter
baumannii for diagnostics and therapeutics
Patent: US 6562958-A 2393 13-MAY-2003;
location/Qualifiers
1..1902
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 25.2%; Score 445.4; DB 6; Length 1902;
Best Local Similarity 55.2%; Pred. No. 4,7e-96;
Matches 932; Conservative 0; Mismatches 746; Indels 9; Gaps 3;

QY 46 GAGCGTCTCGAGCATTTGGGCTAAGACCCGTCAGAACAACTGCGTGTGCGAGGGCG 105
Db 166 GATCGTTTAATTCATTTTGGACAACCAACCGACATATTTTGGACAAACGCAAT 225
QY 106 GCAAAATGGGAATGGCGTGTATCAGTACGCGGAATGTTTCCAAACGTCGCGCATC 165
Db 226 GCTCAAGGCGCAATGGCTCAAACTGAGTTATGCAAGAACTTCTACAAACCGCATGGCAATT 285
QY 166 GCACAGAGCTGTCTTCCTTACGACTATGCGCAGAGGTCCGCTTATGCTCTGGA 225
Db 286 GCTCAAGCTTTGATGCTGCTGTAATTTAAGCCAAAGAACCTTGATCAATTTAAGTGT 345
QY 226 AATGACCTGGAACATCTTCACTGCTGCAATTTGGGCTATGTATGCGGACATTTCCATATGC 285
Db 346 AATGATCTGAACATTTTAACATGCTATGAGTGTGATGCTGAGGCGAGGCTTTCTCG 405
QY 286 CCGGTGCTCTCTGCTTATTAACATGCTGTGCAAGATTTGGCAAGCTGCGTCACATGTA 345
Db 406 GCTATTTTCCCTGCTACTGCTGATTTCTCAAGACTTTGGCAAACTCAAACTATGTGTT 465
QY 346 GGTCTTGTGAACCGGAGCTGTGTGTGCGCGGATGACACCTTTCCACGCGCAATT 405
Db 466 GAAATGCTCACACCTGTATGTATTATGCAAGCATGACGTTTGGCAAGCGATT 525

QY 406 GAGACCATTCCTGCGGACGACGTGCCGAATCTTCACTCGAGCGAATTGACCGGCGG 465
Db 526 CAGGCAATGATTAATGCGCTGATTTGAAGTGTGACCAATTAAGGAATATGCGGCGATCAG 585
QY 466 CGACGAGTAAATTTTGAACAGCTGTGAGACAGCTGTGTGGATTTGAAGGAGATTAATGCC 525
Db 586 ATCTGACAGCTTTTCAATCGCTTATGAATACCAAGTTTCAAAATGTTCAAGAGTTT- 642
QY 526 TTTCGCGCAACTGCGCCCGATATGATTTGCAAGTTCTGTGTTCACTTGTGCTACCAAA 585
Db 643 TATCAAACTTTATGAAAACGATTTGCCAAATTTCTGTTTATCATCAGGTTCAACTAAA 702
QY 586 CTGCTAAGCGGTGCGGACTACTCAACGAAATGCTGTGCGCAATCAAGCATGCTTCTG 645
Db 703 TTACTTAAGCTGTACGACACACATTAATGTGTGTTAATATGACGAAATGTTATTG 762
QY 646 CAAACTTTCCCGGTTTGTGTGAAGCGCGGTGTGTGTGATGATGCTGTGCGGTGGAAC 705
Db 763 CAGACTTTCCTGAGTTTGAAGAAACACCGCTGTCTACTCGACTGTGCTGTCTTGGCAC 822
QY 706 CACACCTTGGCGGCGGACGCAACATCGGATGCTGTGTGTAACAAGCGGACGTAATAC 765
Db 823 CACACATTTGGCGGAGTCAATGTGCGCATGCACTCTAATAAGCGGTACGATTTAC 882
QY 766 CTTGACGACGCTTAACCAACCGCCCAAGGTTGCGGAGACGCTTGCAACTTGAGCGAA 825
Db 883 ATGATGATGAGCAAAACCGTTGAGGAAATTTGACGAACTATGTGTAATCTCAAGAA 942
QY 826 ATCTCTCCACGCTGCTACTGTGCGGAAAGCTGTGGAAGAAATTAAGTGGTGGCTT 885
Db 943 ATTTCTCAACTGTATTATTAATGTGCCAAAGTTGGGAACACTCAACGAAAGGTTA 1002
QY 886 GAGCGAGACGATCCCTGCGCGCAACGCTTCTGCTGCTGATGAAGCTGTCTTCTGCGG 945
Db 1003 GAAAAATGATGAATTTAAGAGCGCTTTTGTCCAAAGTTAAATTTATCTTTGCC 1062
QY 946 GCGGCTGGGTGTGCGAAGGATCTGGAATGTTTGTGACCCGGTGTGTAACGACACTGT 1005
Db 1063 GGTGTGCACTTTCAAGAGGCGGTGGAACAGACTGATTAATATGCTCTGACAACTTGC 1122
QY 1006 GGTGACGCAATTTGCGATATGCGGGGTGTGCGGATACGGAACGTCCTCTTCTGCACT 1065
Db 1123 GGAAGAAAAATCGCATTTATGAGCGGATGAGGATACGGAACGTCCTCTTGTGCT 1182
QY 1066 TTTAACACCGGACCGCTGCTGATGCTGTGTTAATTTAGTGGCTGCGACGCTGTGCGAG 1125
Db 1183 TTTACAACTGCGCCACGCGTAATGCTGTGCTTTATTTGTTAATCCGTCTCCGGATGGA 1242
QY 1126 GTCAAGCTGTGTCGCTGATGAGGAATTTGAAAGGCGTTTCCATGTGCGCACGTCATG 1185
Db 1243 ATTAAGCTAGTTCAATGTGTGACAACTTGAATTTTGTGCTGTGCGCAACGTCATG 1302
QY 1186 AGGCGCTACTGGG---TGCTCTGAACAAAAATGCCAAGCGTTGACAGGAAAGGCTAT 1242
Db 1303 AAAGGCTATTGGGCTTAAAGGCGAACCAAAAGTACTAATTTTATATGATGAAGGCTTT 1362
QY 1243 TACTGCTCCGCTATGCTCAATTAATTTGAGCATCTGCGGATCTCGAAGGCTGTGATG 1302
Db 1363 TTTCATACAGGCGATGCTGCTGTATGTCATGTCAATGATCAACAAAGGCTTAATG 1422
QY 1303 TTGACGCTGCAATTTGCTGAAGCTTCAAGCTGTCTCAAGGAGTATTTGTCAAGCTTGGG 1362
Db 1423 TACGAGCGAAGATTTCCGAAGCTTTAACTCAATACAGGCACTTTTGTGATATGTCGGC 1482
QY 1363 CCAATGCGACGCGGCGGTCTGGAAGCGGCTTTAAGTCTCTGACGACGATGATGTGCT 1422
Db 1483 ACACTACGCAACAAATGCTTATTTCAAGGTAATTTACTATCCAAATGTTGTATTAAT 1542
QY 1423 GCTCTGATGCTGAATGCTTGTGATGCTGTGTGCTGCTGTGCTGTGCTGCTGCTGCTGCT 1482
Db 1543 GGTTCAAACCTGAATGCTATTTGTTTCTGATTTTTCAAAAATTAACGCTTGTGCTCAA 1602

Qy	1483	TTTGCGGGGCTAGGAAAAAGAGCG---TCGGAGCCGAGGTGCTTGCCAGTAGCCGGTT	1539	MEDLINE	94341565
				PUBMED	8063101
Db	1603	TATCAGAGCTCTTAAGCTGGGGGAATATTTCTGACAGAGATACCTGCAACACCTTAAATC	1662	REFERENCE	6 (bases 1 to 45324)
Qy	1540	CGGGCCCTGTTTGTCTGACTGGCTCAAAAGACTCAATCGAGAACTGGCAATGCCAGT	1599	AUTHORS	Elismore,D.A. and Ornston,L.N.
				TITLE	The pca-pob supraoperonic cluster of Acinetobacter calcoaceticus contains quiaA, the structural gene for quinate-shikimate dehydrogenase
Db	1663	CAACAAATGGTTCCCGCAATTTTAAACGACTTTAATAAATGCGATGCGAGCTCAAAAT	1722	JOURNAL	J. Bacteriol. 176 (24), 7659-7666 (1994)
Qy	1600	CGCATCATGTGGGTAGGGCTCTCGATACGCCCGCTCGATTTGATTAAGGCGAGTCACT	1659	MEDLINE	95095936
				PUBMED	8002591
Db	1723	ACAGTCTCAATGCTTTATTTAATACCGAGACCTCAGTTAGATGCCGGGAAGTGACC	1782	REFERENCE	7 (bases 23320 to 27131)
Qy	1660	GACAAGGAGCTCGATCAACGAGCGCTGTTTGGCAATGCGCGCAAGTGGATGCGC	1719	AUTHORS	Elismore,D.A. and Ornston,L.N.
				TITLE	Unusual ancestry of dehydratases associated with quinate catabolism in Acinetobacter calcoaceticus
Db	1783	GATTAAGGCAACCTCAATCAAGAAGTATTACAAAGCTCGTCCGCTTTAATTGATGAG	1842	JOURNAL	J. Bacteriol. 177 (20), 5971-5978 (1995)
Qy	1720	CTGTATC	1726	MEDLINE	96011389
				PUBMED	7592351
Db	1843	CTTTATC	1849	REFERENCE	8 (bases 1 to 45324)
				AUTHORS	Gerischer,U., Segura,A. and Ornston,L.N.
				TITLE	pcaU, a transcriptional activator of genes for protocatechuate utilization in Acinetobacter
				JOURNAL	U. Bacteriol. 180 (6), 1512-1524 (1998)
				MEDLINE	98175676
				PUBMED	9515921
				REFERENCE	9 (bases 1 to 9706)
				AUTHORS	Parke,D., Garcia,M.A. and Ornston,L.N.
				TITLE	Cloning and genetic characterization of dca genes required for beta-oxidation of straight-chain dicarboxylic acids in Acinetobacter sp. strain ADP1
				JOURNAL	Appl. Environ. Microbiol. 67 (10), 4817-4827 (2001)
				MEDLINE	21455042
				PUBMED	11571189
				REFERENCE	10 (bases 1 to 45324)
				AUTHORS	Parke,D. and Ornston,L.N.
				TITLE	Hydroxycinnamate (hca) catabolic genes from Acinetobacter sp. strain ADP1 are repressed by HcaR and induced by hydroxycinnamoyl-CoA thioesters
				JOURNAL	Appl. Environ. Microbiol. (2003) In press
				REFERENCE	11 (bases 1 to 45324)
				AUTHORS	Smith,M.A., Weaver,V.B., Young,D.M. and Ornston,L.N.
				TITLE	Genes for chlorogenate and hydroxycinnamate catabolism (hca) are linked to functionally related genes in the dca-pca-quI-pob-hca chromosomal cluster of Acinetobacter sp. strain ADP1
				JOURNAL	Appl. Environ. Microbiol. 69 (1), 524-532 (2003)
				MEDLINE	22401456
				PUBMED	12514037
				REFERENCE	12 (bases 12470 to 15266)
				AUTHORS	Gerischer,U.C.
				TITLE	Direct Submission
				JOURNAL	Submitted (15-DEC-1993) Biology, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8103, USA
				REFERENCE	13 (bases 27214 to 29733)
				AUTHORS	Elismore,D.A.
				TITLE	Direct Submission
				JOURNAL	Submitted (28-JUN-1994) Biology, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8103, USA
				REFERENCE	14 (bases 23320 to 27131)
				AUTHORS	Elismore,D.A.
				TITLE	Direct Submission
				JOURNAL	Submitted (25-JAN-1995) Biology, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8103, USA
				REFERENCE	15 (bases 1 to 45324)
				AUTHORS	Ornston,L.N.
				TITLE	Direct Submission
				JOURNAL	Submitted (02-JUN-1998) Biology, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8103, USA
				REFERENCE	16 (bases 1 to 45324)
				AUTHORS	Ornston,L.N.
				TITLE	Direct Submission
				JOURNAL	Submitted (20-FEB-2001) Biology, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8103, USA
				REMARK	Sequence update by submitter
				JOURNAL	
RESULT 12					
ACCP/CAOP/c					
LOCUS	ACCP/CAOP	45324 bp	DNA	linear	BCT 27-JUN-2003
DEFINITION	Acinetobacter sp. ADP1 dca-pca-quI-pob-hca supraoperonic gene cluster, complete sequence.				
ACCESSION	U055770	AT056039	U03407	U13114	M33798 U04359 U11554 U20284
VERSION	U055770.5				GI:32306886
KEYWORDS					
SOURCE	Acinetobacter sp. ADP1				
ORGANISM	Acinetobacter sp. ADP1				
	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter.				
REFERENCE					
AUTHORS	1 (bases 1 to 45324)				
	Hartnett,C., Neidle,E.L., Ngai,K.L. and Ornston,L.N.				
TITLE	DNA sequences of genes encoding Acinetobacter calcoaceticus protocatechuate 3,4-dioxygenase: evidence indicating shuffling of genes and of DNA sequences within genes during their evolutionary divergence				
JOURNAL	U. Bacteriol. 172 (2), 956-966 (1990)				
MEDLINE	90130333				
PUBMED	2298704				
REFERENCE					
AUTHORS	2 (bases 1 to 45324)				
	DiMarco,A.A., Averhoff,B.A., Kim,E.E. and Ornston,L.N.				
TITLE	Evolutionary divergence of pobA, the structural gene encoding p-hydroxybenzoate hydroxylase in an Acinetobacter calcoaceticus strain well-suited for genetic analysis				
JOURNAL	Gene 125 (1), 25-33 (1993)				
MEDLINE	93194074				
PUBMED	8449410				
REFERENCE					
AUTHORS	3 (bases 1 to 45324)				
	DiMarco,A.A., Averhoff,B. and Ornston,L.N.				
TITLE	Identification of the transcriptional activator pobR and characterization of its role in the expression of pobA, the structural gene for p-hydroxybenzoate hydroxylase in Acinetobacter calcoaceticus				
JOURNAL	U. Bacteriol. 175 (14), 4499-4506 (1993)				
MEDLINE	93322329				
PUBMED	8331077				
REFERENCE					
AUTHORS	4 (bases 1 to 45324)				
	Hartnett,G.B. and Ornston,L.N.				
TITLE	Acquisition of apparent DNA slippage structures during extensive evolutionary divergence of pad and cad genes encoding identical catalytic activities in Acinetobacter calcoaceticus				
JOURNAL	Gene 142 (1), 23-29 (1994)				
MEDLINE	94237485				
PUBMED	8181753				
REFERENCE					
AUTHORS	5 (bases 1 to 45324)				
	Kowalchuk,G.A., Hartnett,G.B., Benson,A., Houghton,J.E., Ngai,K.L. and Ornston,L.N.				
TITLE	Contrasting patterns of evolutionary divergence within the Acinetobacter calcoaceticus pca operon				
JOURNAL	Gene 146 (1), 23-30 (1994)				

REFERENCE 17 (bases 1 to 45324)
AUTHORS Smith,M.A., Young,D.M. and Ornston,L.N.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2001) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
18 (bases 1 to 45324)
AUTHORS Smith,M.A., Young,D.M. and Ornston,L.N.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-2002) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
REMARK Sequence update by submitter
REFERENCE 19 (bases 1 to 45324)
AUTHORS Parke,D. and Ornston,L.N.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2003) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
REMARK Sequence update by submitter
COMMENT On Jun 27, 2003 this sequence version replaced gi:20492839.
FEATURES
source
1. .45324
/organism="Acinetobacter sp. ADP1"
/mol_type="genomic DNA"
/strain="ADP1"
/db_xref="taxon:62977"
/note="Formerly Acinetobacter calcoaceticus"
/gene="dcaF"
complement (<1. .689)
/gene="dcaF"
/function="putative thiolase in catabolism of dicarboxylic acids"
/codon_start=1
/transl_table=1
/product="DcaF"
/protein_id="FAL09090.1"
/db_xref="GI:15812038"
/translation="MLNAVLYDGLRTPFGHAGELASIRPDDIAGLVIOQLIEKTGVA
GADIEDVTFEDTQAGDSNRIRAHALAGLPVPGQVNRILCASGLAAIIDSARA
ITCGEBDLVJAGVSWSRAPFVWGAESYSDAKIYDTTIGTFENKKIIVQYGGH
SMPTGDNVAVEYGISREQADLPAAQAKYQALAEFGFAGETITAVEGSKKLPRK
QVTEDEHPRPS"
complement (710. .2227)
/gene="dcaH"
complement (710. .2227)
/function="hydroxyacyl-CoA dehydrogenase in catabolism of dicarboxylic acids"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="DcaH"
/protein_id="FAL09091.1"
/db_xref="GI:15812039"
/translation="MTHPIKKAIIIGVWVGSGIAQIAQSGHITVLYDAKAAQQA
KQDLATPQKLDKNKITTAYDAANANLILANEHLDCDLIVEIVERLIDKSL
MSQLEAIVPETTILASNTSLSTIALASNKPERVAGYFNPVPLMKVVEIIGKSL
TDPKHEITLNLQSLRVLGHRPVVAKDTPGFIINHAGRAYEALAKIENETDSEIDR
ILDDGFGFPMGLMDLTGLDVSHPMESIYHOYEBAAYRNSLTAKOMLAKROK
KVGQGFYRTSGTGETSAKVAERLLTVPKWLIADFEEDKOLLINLYTHNIOIDY
GAKPOADSLTACVGEDTHALRIANVPANHSVALIDMLYGIKERTIMSLTBTY
SHAHSITFLNDGAMSTTIGSIFVAKRIILAVINAGCDIAQAIAISVDINAVRIG
LGYFPGPIEWGDISNKLILLNRIALTISDRYRSPWLQRRVALNPLPTT"
complement (2276. .3037)
/gene="dcaC"
complement (2276. .3037)
/gene="dcaC"
/note="putative ketoacyl-CoA reductase"
/codon_start=1
/transl_table=1
/product="DcaC"
/protein_id="FAL09092.1"
/db_xref="GI:15812040"
/translation="MAKVLITAGSGIGRICIAEAPLKQGDYFVCDINDSHQPRQ

Query Match 25.0%; Score 442.8; DB 1; Length 45324;
Best Local Similarity 55.2%; Pred. No. 2e-95;
Matches 930; Conservative 0; Mismatches 747; Indels 9; Gaps 3;
DYPQLHTFCCDLGQPRYVQAMPDTAVOOLGDLILNNTGVSGPTVAEELSPFEWTO
VNLINNSTFLVTKLAIPYLKQSDAGVILNISIAARMGYPFLAYSTSKGIWGFTR
TISEMGADIRVNALIPGAVDERVORVQAADSMNISIEDVTENALANQSIKHFV
NPKHIADLCIFLASDSGRISQIILPIDDKRLT"
46 GAGGCTCGAGCATTTGGCGTAGACCCGTCAGAAACCAACCTGCGTGGCAGGCG 105
Db 39842 GATCGTTGATTCATTTTGTCTAAAGTTAGCCCTCACATACCTTTGCGCTTAAACGCAAT 39783
Qy 106 GCAATGGGGAATGAGCGTCGTATCAGCTACGCGGAAATGTTCCACACATCGCGGCATC 165
Db 39782 ACAAGAGCAATGATGATCAGGCTCAGTTATGAGAAACCTTGCAAGTGTGGCATAT 39723
Qy 166 GCAAGAGCTTGTCTTCTTACGAGCTATCGGACGAGCGTCCGTCTTATCGTCTGGA 225
Db 39722 GCAAGAGCTTGTCTTACGAGCTATCGGACGAGCGTCCGTCTTATCGTCTGGA 39663
Qy 226 AATGACCTGGAATCATCTCAGCTGCGATTTGGGGCTATGATGCGGGCATTCCTATTCG 285
Db 39662 AATGATCTTGAACATTTAACGCTTTGATGAGGCGCAATGTTGGCAGGTGATACATTTTCT 39603
Qy 286 CCGGTGCTCTGCTTATTCATCTGCTGCGCAAGATTTGGCGAAGCTGCGTCAATCTGA 345
Db 39602 GCTATTTTGGCAGCGTACTACGATATCAAGATTTTGGCAATTTAAACATGTTT 39543
Qy 346 GGTCTTCTGCAACCGGAGCTGCTTCTTGTGCGGATGACGACCTTTCCAGCGGCAAT 405
Db 39542 GATGTGCTACCCGACGAGTGGTGTATGCAATGATGACCAACCTTTTGTGTAAGGCAT 39483
Qy 406 GAGACCATTCGTCGCGGACGACGCGGCAATCTTCACTGAGGCAATGGCGGCGG 465
Db 39482 CAAGCTGTACGCAAGATCATATTGAATTTGTACCAATTAAGCATTTATAGTGATCA 39423
Qy 466 CGCAGGAGTATTTTGAACAGCTGCTGAGACAGCTGTGGGATGAGGACAGATTAATGCC 525
Db 39422 GTCTGTACATCGTTTCACTTATTAATGATACCCGTGACCGATGTTCAAGAACATTA 39363
Qy 526 TTTGGCGCACTGCGCCCGCATATGATTTGCCAATTTCTTTTCACTTGTGCTCTACCAA 585
Db 39362 GAAACCTGATATGAGC--ATCAGATTTGCCAATTTTATTCATTCGCGTTGACCAAG 39306
Qy 586 CTGCTTAAGCGGTGCGCACTACTACGCAATGCTGCGGCAATGACAGATGCTTCTG 645
Db 39305 CTTCGAAGGCTTTTCCGACACACATTTGATGCTGTATCAATCAACAAATGCTACTT 39246
Qy 646 CAACCTTTCCGCTTTTGTGTAAGACCGCGGTGCTGTGTGCTGCTGCGTGGAC 705
Db 39245 CAACATTTCCAGAAATTTGAAGACACACCGCGCTGTCTTACTGACGTGTGCTTGGCAT 39186
Qy 706 CAACCTTGGCGGCGGACCCCAACATCGGATGCTGTGTGTAACAGCGCGGACGTACTAC 765
Db 39185 CATACATTCGTGTGAAGCCCAATGTTGGATGCGCTTTATTAACGCGGAAACATTTAC 39126
Qy 766 CTGAGAGCGGTAAACCAACCGCCCAAGGTTGCGCGAAGCGCTTGCAACTGAGCGAA 825
Db 39125 ATGATGACGATAGCCCGTTCGGGAAAATTTGATGAACGATTTGAAATTTAAAGAA 39066
Qy 826 ATCTTCCCATGCGTACTCACTGTGCGGAAAGGCTGGAGAGATTAAGTGGTGCCTT 885
Db 39065 ATTTCAACGACGCTTTATCTAATGTACCAAAAGGCTGGGAAAGATCAACCGATGACCTT 39006
Qy 886 GAGCGAGACATACCTCGGCGCAACGCTTCTTGTGCTGATGAAGCTGTTCTTTCGGC 945
Db 39005 GAAAGAGATGCCAATCCAGAGATCGCTTCTTGGCAACGTTAAATTTGTTTGTGCT 38946
Qy 946 GCGGCTGGGTGTGCGAAGGATCTGGAGATGTTTGAACCGGCTGCTGAACGACTGT 1005
Db 38945 GGTGCTGATTTATCTGAAGAGGCTGGAACGACTGATTAATTTGACAGGACATTTGC 38886

Qy	1006	GGTAGCGCATTCGATGATGCGGCTCTGGGCAATGACGGAGACTGCTCTTCCGCACT	1065
Db	38885	GGTGAATAAATTCGATCATGATGCTTGGGCAATGACGGAGACTGCTCTTCCGCACT	38826
Qy	1066	TTTACACCGGACCGCTGTGATGCTGTGTTCAATTGGGCTGCGAGCGCTGCGAG	1125
Db	38825	TTTACACCAAGCCCGGATGATGCTGTGTTTATTTGCTACCTGCTCCGGGCTGTGAA	38766
Qy	1126	GTCAGCTCGTTCGCTGCTGATGAGAAATTGGAAGGCGTTTCATGCTCCGACATG	1185
Db	38765	ATCAAAATTAATTCCTTATGATGACAACTGAAATTTGTGTTCTGTGCAAGCACTATG	38706
Qy	1186	AGGGGCTACTGGGCG---TGCTCTGAAACAAATGCGCCAGCGGTGCAAGAGGCTAT	1242
Db	38705	AAAGGTTACTGGCGCCCTTAAGGCCGACCAAGCAACATTTTGTATGAAGACGGCTTT	38646
Qy	1243	TACTGCTCGGCTGATGCTCAATTAATGACAGATCTGCGATCTCAGAAAGTCTGATG	1302
Db	38645	TTCGATCTGCGGATGCGGTACGTCTGTGATCAAAAGCACTGCAAGAGGCTGATG	38586
Qy	1303	TTTGAAGCTGGAATTCCTGTAAGACTTCAAGCTGCTCAGGGGATTTTGTACAGCTGGG	1362
Db	38585	TACGACGCGCGTATTTGCAAGATTTCAAACTCAACGCGGTACCTTTGTGAATGTGGC	38526
Qy	1363	CCATTCGCAAGCGGCGGCTTCTGAAAGCGGCTCTTACGCTCTGACGTAGTGTGCT	1422
Db	38525	ACTTTACGCAACAAAGGACATTCATTAAGGTATTTGTGATTCAGAAATGTTGTATTAAC	38466
Qy	1423	GCTCTGATGCTGATGCTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1482
Db	38465	GGTTCAATCTAAATGCGATTTGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	38406
Qy	1483	TTGTGCGGGCTGAGAAAGAGGCGTCCGAGCGCGAGGTGCTGCGAGTGAAGCGG---	1539
Db	38405	TTTGTGCTGTGATGATATGAAATATCATAGCGCAATGATATCTTGCAACACCTTAAAGTA	38346
Qy	1540	CGGCGCTGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1599
Db	38345	CAGCAGTGTGTTTGCACAGTTTATTTATGATGACAAAGATGCGGCGGAGTTCAAAT	38286
Qy	1600	CGCATCATGTTGGGATGAGGCTCTGCTGATACCGCGCGCTGATGTAAGGCGGAGTCACT	1659
Db	38285	CGGGTATCAATGTTGATTTAATGATGAGACACACAGCTGATGAGGTAAGCAAC	38226
Qy	1660	GACAGGCGCTGATCAACAGCGCGCTGTTTTCATGAGGCGGCTGCGGAAATGATGCG	1719
Db	38225	GATTAAGGCAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG	38166
Qy	1720	CTGTAT 1725	
Db	38165	CTGTAT 38160	

RESULT 13
CR543861.17/c
Sequence split into 36 fragments LOCUS CR543861 Accession CR543861

Fragment Name	Begin	End
CR543861_00	1	110000
CR543861_01	100001	210000
CR543861_02	200001	310000
CR543861_03	300001	410000
CR543861_04	400001	510000
CR543861_05	500001	610000
CR543861_06	600001	710000
CR543861_07	700001	810000
CR543861_08	800001	910000
CR543861_09	900001	1010000
CR543861_10	1000001	1110000
CR543861_11	1100001	1210000
CR543861_12	1200001	1310000
CR543861_13	1300001	1410000
CR543861_14	1400001	1510000

Qy	CR543861_15	1500001	1610000
Db	CR543861_16	1600001	1710000
Qy	CR543861_17	1700001	1810000
Db	CR543861_18	1800001	1910000
Qy	CR543861_19	1900001	2010000
Db	CR543861_20	2000001	2110000
Qy	CR543861_21	2100001	2210000
Db	CR543861_22	2200001	2310000
Qy	CR543861_23	2300001	2410000
Db	CR543861_24	2400001	2510000
Qy	CR543861_25	2500001	2610000
Db	CR543861_26	2600001	2710000
Qy	CR543861_27	2700001	2810000
Db	CR543861_28	2800001	2910000
Qy	CR543861_29	2900001	3010000
Db	CR543861_30	3000001	3110000
Qy	CR543861_31	3100001	3210000
Db	CR543861_32	3200001	3310000
Qy	CR543861_33	3300001	3410000
Db	CR543861_34	3400001	3510000
Qy	CR543861_35	3500001	3598621

Continuation (18 of 36) of CR543861 from base 1700001 (CR543861 Acinetobacter sp. ADP1 c

Qy	46	GAGGCTTCGACGATTTGGGCTTAAGACCCGTCAGAACAAACCTGCTGCTCCAGGCGC	105
Db	33693	GATGCTTGAATTCATTTTGTCTAAAGTTAACCTTACCTATCTTTGCGCTTAACGCAAT	33634
Qy	106	GCAATGGGGAATGGCGCTGATCAGCTACGCGGAATGTTCCACAAAGTCCGCGCATC	165
Db	33633	ACACAGACAAATGATGATCAGGCTCAGTTATGCAAGAACTTGCACAGCTGCGCAATTT	33574
Qy	166	GCAAGAGCTTGTCTTCCCTTAACGATACGCGAGAGCGTCCGCTTATGCTCTGGA	225
Db	33573	GACAAAGCTTTGATCAGCGCACTTAAGCCAAAGAACGACTTTGTATTTATCGGGA	33514
Qy	226	AATGACCTGGAACATCTTCACTGCTGCACTTTGGGCTATGATGCGGCAATTCCTATTGC	285
Db	33513	AATGATCTGAAACATTAACGCTTTCATGAGGCGGATGTTGGCAGGTGATCAATTTCT	33454
Qy	286	CGGCTGCTCTGCTTATTAACGCTGCTGCAAGATTTGGCGAAGCGGCTCAATGTA	345
Db	33453	GCTATTTGCGCAGCTACTGATGATCTCAAGATTTGGCAATTAAGCATGCTTT	33394
Qy	346	GGTCTTCTGCAACCGGAGCTGCTTGTGCTGCGATGCAACACTTTCCAGCGCAATT	405
Db	33393	GATGTGCTACCCGAGCATGATGATGCAATGATGACAGGATTTTGGTAAAGCAATT	33334
Qy	406	GAGACATTTGCGCGGAGCGAGCTGCGGCAATCTTCACTGAGGCGAATTTGGCGGCGG	465
Db	33333	CAAAAGCTGTACGCAAGATTCATATGAAATTTGATCAACATTAAGCATTAATAGGATCA	33274
Qy	466	CGACGCTGAGTTTGTGACCAAGCTGCTGAGACAGCTGTGGAGATGAGGAGATTAATGCC	525
Db	33273	GTCGTACATGCTTTCAGTCTTATTAATGATTAACCCCTGTGACCGATGTTCAAGAACATTAT	33214
Qy	526	TTTGGCGCAACTGGCCCGATGATGATGCAAGTTCTTGTTCATCTTGTGCTGTACCAA	585
Db	33213	GAAAGCTAATATGAGC---ATCAGATTTGCAAGTTTATTTATCATCGGTTGACCAAG	33157
Qy	586	CTGCTTAAGCGGTTGCGGATCACTACAGCAATGCTGCGGCAATCAGCAAGATGCTTCTG	645
Db	33156	CTTCCAAAGGCTGTTCGACCAACATTTGATGCTGTATCATCAACAAATGCTACTT	33097
Qy	646	CAAACTTCCGCTTTTGTGTAAGAGCGCGGCTGTGATGATGCTGCTGCGGTGCAAC	705
Db	33096	CAAACTTTCAGATTTGAAGACACCGGCTGTCTTACTGAGCTGTGTCTTGTGCAAT	33037
Qy	706	CACACTTTCGCGGACCAACATCGGATGCTGTGTATCAACGCGGACGCTACTAC	765

Db 33036 CATACATTGCGTGAAGCAATGTTGGATTGCTTTATTAACGGCCGACCATTTAC 32977
QY 766 CTGACAGCGGTAAACCAACCGCCCAAGGGTTGCGCGAGACGCTTGCAACTTGAGGAA 825
Db 32976 ATCGATGAGCGGTAGCCCGTTCCGGGAAAATTGATGAAAACGATTGAAAATTAAAGAA 32917
QY 826 ATCTCTCCACCTGCGTACCTCACTGTGCGGAAAAGGCTGGAGGAATTAAGTGGTGCCTT 885
Db 32916 ATTTACACGACGGTTTATCTTAATGTACCMAAGGCTGGGAGAACTCACCGATGACCTT 32857
QY 886 GAGCGAGACATACCTCGGCGGAAACGCTTCTTGCTGCGATGAAGCTGTTCTTCCGCG 945
Db 32856 GAAAAAAGATGCCAATCAGAGATCCCTCTTGCCAAACGTTAAATTTTGTTTTTGCT 32797
QY 946 GCGGCTGGGTTGTCGCAAGGATCTGAGATCTGTTGACCGGGTCCGTGAAACGACACTGT 1005
Db 32796 GGTGCTGCAATATCTGAGACAGGCTGMAACCGACTGATTAATTTGACAGGACATTTGC 32737
QY 1006 GGTGACCGCATTCGATGATGCGGGTCTGGGCAATGACGAGACTGCTCTTCTGCACT 1065
Db 32736 GGTGAAAAAATCCGATCATGATGCGCTTGCGCATGACCGAAACGGCACCATGCTGCGCC 32677
QY 1066 TTATACACCGGACCGCTGTCATGCTGCTGTTACATTGGGCTGCGACGCTGCTGCGAG 1125
Db 32676 TTATACACAGGCCCAAGGTAATGCGCTGTTTATTTGGCTAACCTGCTCCGGGCTGTGA 32617
QY 1126 GTCAAGCTCGTTCCGGTCAATGGAAGGAGGCTTCCATGGTCCGACGTCATG 1185
Db 32616 ATCAATTAATGTTCCCTATGCTGATGCAAACTTGAATTTTGTTGTTGTCGCAAGCACTGATG 32557
QY 1186 AGCGGCTACTGCGG---TGTCTGTGAACAAATGCCCAGCGTTGACGAGAGGCTAT 1242
Db 32556 AAGGTTACTGCGCGCTTAAGCGCCGACACGACAAACAAATTTTGATGMAAGCGGCTTT 32497
QY 1243 TACTGCTCCGCTGATGCAATTAATGCGAGATCTGCGCATCTCAGAAAGCTGCAATG 1302
Db 32496 TTTCATATCTGCGCATGCGGATGCTGTGTCGATCAAAACGACCTCGACAGGGGCTGATG 32437
QY 1303 TTATGCGGCTGCAATTCCTGAAGACTTCAAGCTGCTCAGGGGTAATTTGTCAGCGTTGGG 1362
Db 32436 TACGACGCGCGTATTCGAAAGATTTTCAAACTCAACACGGGTACCTTTGGAATGTGGG 32377
QY 1363 CCATTGCGCACGCGGCGGCTTGAAGCGGCGCTTTACGTCCTGACGTAAGTGTGCT 1422
Db 32376 ACTTTACGCAACAAAGCACATCATTCAGGTAATTTGTCATGATGATTTGTATTACC 32317
QY 1423 GCTCCGATGCTGAATGCTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1482
Db 32316 GGTTCAAATCTTAATGCGATTGGCTTGTGTTTCTTAAGCTTGAATGCTGCTGCAAT 32257
QY 1483 TTGTCCGGGCTAGAAAAGAGCGGTGSAAGCGCGAGGTGCTTGCACATGAGCCG---GTT 1539
Db 32256 TTGTGCTGCTGATTAATGAAAATCATAGGCAATAGATATCTTGCAACACCTTAAGTA 32197
QY 1540 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1599
Db 32196 CAGCAGTGTGCTTGAAGATTTTAATTCAGTACAAAGATGCCACGCGGCAATTCAAAT 32137
QY 1600 CGATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1659
Db 32136 CCGGATTCATATGTTATTAATGACTGAAGCACCAAGCTCGATGAGGTGAGTAACC 32077
QY 1660 GACAAAGGCTGATCAACACGCGCTGTTTGTGCAATGCGGCTGCGGAAAGTTGATGCG 1719
Db 32076 GATTAAGGCAATTTAATGAGATTAATCTTAAACGTCGTCGCACTGCTGCAAGAA 32017
QY 1720 CTGTAT 1725
Db 32016 CTGTAT 32011

RESULT 14
AX024534

LOCUS AX024534 1203 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 45 from Patent DE19850242.
ACCESSION AX024534
VERSION AX024534.1 GI:10184691
KEYWORDS
SOURCE
ORGANISM
Pseudomonas sp. HR199
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 Overhage, J., Priefert, H., Rabenhorst, J. and Steinhuechel, A.
Patent: DE 19850242-A 45 04-MAY-2000;
HAARMANN & REIMER GMBH (DE)
FEATURES
source
1..1203
/organism="Pseudomonas sp. HR199"
/mol_type="unassigned DNA"
/db_xref="taxon:86003"
342..>740
/note="unnamed protein product; N-Terminus (AMINO ACID 1
bis 133) codierender Teil des
Ferulasaure-CoA-Synthetase-Gens (fcs) "
/codon_start=1
/transl_table=11
/protein_id="CAC08976.1"
/db_xref="GI:10184692"
/translation="MRSLEALPLPGRILIERLEHNAKTPEQTCVAARANGEMRRIS
VAEMFNVDALIAISLLPYGLSABRPILIVSGDLEHLQAFGAMVAGIPIYCPVSPAYS
ILSODLAKIRHIVGLQPLVFPADAAPQR"
<741..827
/note="unnamed protein product; C-Terminus (AMINO ACID 562
bis 589) codierender Teil des
Ferulasaure-CoA-Synthetase-Gens (fcs) "
/codon_start=1
/transl_table=11
/protein_id="CAC08977.1"
/db_xref="GI:10184693"
/translation="AVLQWRSKVDALYKGEDQSLMRDEATL"
ORIGIN
Query Match 22.9%; Score 404.6; DB 6; Length 1203;
Best Local Similarity 91.6%; Pred. No. 3e-86;
Matches 428; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 ATGCGTTCTCTGAGGCGCTTCCCTCCGCGGTGCAATCTTGAGGCTCGAGCAT 60
Db 342 ATGCGTTCTCTGAGGCGCTTCCCTCCGCGGTGCAATCTTGAGGCTCGAGCAT 401
QY 61 TGGGCTAAGACCGCTTCAAGCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Db 402 TGGGCTAAGACCGCTTCAAGCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 461
QY 121 CGTCGATAGAGCTAGCGGGAATGTTCAACAAGTCCGCGCCCATGCGACAGAGCTTGTT 180
Db 462 CGTCGATAGAGCTAGCGGGAATGTTCAACAAGTCCGCGCCCATGCGACAGAGCTTGTT 521
QY 181 CCTTAAGCATATCGGCAAGGCTCCGCTGCTTATGCTGCTGGAATGACCTGGAACAT 240
Db 522 CCTTAAGCATATCGGCAAGGCTCCGCTGCTTATGCTGCTGGAATGACCTGGAACAT 581
QY 241 CTTACAGCTGCAATTTGGGCTATGTAATGCGGCAATTCCTATTGCCGGTGTCTCTGCT 300
Db 582 CTTACAGCTGCAATTTGGGCTATGTAATGCGGCAATTCCTATTGCCGGTGTCTCTGCT 641
QY 301 TATTGACCTGCTGCGAATTTGGGAGCTGCGGCAATGATGATGATGATGATGATGATGATGAT 360
Db 642 TATTGACCTGCTGCGAATTTGGGAGCTGCGGCAATGATGATGATGATGATGATGATGATGAT 701
QY 361 GGAAGTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 702 GGAAGTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 761
QY 421 GAGCAGTCCCGCAATTTTCACTCGAGGCGAATTTGGCGGCGCGG 467

Db 762 GCAGAGTTGATGCGCTGATCGGTGATGAGATCATCATCTGCGC 808

RESULT 15
AE008067/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ORGANISM

AE008067 10173 bp DNA linear BCT 28-MAY-2004
Agrobacterium tumefaciens str. C58 circular chromosome, section 125
of 254 of the complete sequence.
AE008067 AE007069
AE008067.1 GI:15156484

Agrobacterium tumefaciens str. C58
Agrobacterium tumefaciens str. C58
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
1 (bases 1 to 10173)
Hinkle,G., Slater,S.C. and Goodner,B.
Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
disease in Plants
Unpublished
2 (bases 1 to 10173)
Hinkle,G., Slater,S.C. and Goodner,B.
Direct Submission
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA

FEATURES
source
1..10173
/organism="Agrobacterium tumefaciens str. C58"
/mol_type="genomic DNA"
/strain="C58"
/isolate="Cereon"
/db_xref="taxon:176299"
/complement (100..1263)
/gene="AGR_C_2609"
/complement (100..1263)
/gene="AGR_C_2609"
/note="branched-chain amino acid ABC transporter,
periplasmic amino acid-binding protein - Deinococcus
radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_C_2609p"
/protein_id="AAK87205.1"
/db_xref="GI:15156485"
/translation="MGKTMKRIINTWMAVALTTAGSAFADVTIGVITPTTGGRA
ALGIPLKOSIEFWPADIDEGKLVILLDDAGDPSVATTNRRASDDKVDVIGSALT
PAIAVAGVANETEVHPFALSPITPAATKKNFMVQSFVLVKSMEFARMQKXGRV
AFGPAFDSYGQGVADALAKATGKXGVLVAEERYARADSVSGOVLKSLAAPDAIFI
GASGTGAALPOIVRRRGKPKKRYVVSAGVTPPLIAGKADNVTAFASGPVVAEDQ
PDSAPTKQEGSLATLKAPEGKFGENTRIQFSGHWVDWVILKAVPALKKAKPGTAEF
RAALRDAIETSGPVNATHGYEFTPAHNHGLDORGVLLTPAADGKFFVLIDNK"

gene
CDS
/gene="AGR_C_2611"
/complement (1317..2960)
/gene="AGR_C_2611"
/note="probable acyl-CoA dehydrogenase - Deinococcus
radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_C_2611p"
/protein_id="AAK87206.1"
/db_xref="GI:15156486"
/translation="MKSQGMVASASRGDVNSFTQAAATMELLKATSGMRRIOGLR
GICDDALVAAILNEARFAETVLEPLATPDREGCTVVDGVRVPTGPADEFRAHAGD
GMIAVDVPERFGGALPLVQAACAPLFEFGCVAAVMASGSTRACHALAEETGAHVA
DEVVPRILAGEMAATTCISRPDAGSDLARLSAARSGAMTEGRKIMTSRQDHMA
MRGHCLARTGETRTGSLPLVDEIDENKGYTVERIEKMGILHSPITALFD
GAGLLGEGRGALQIPAVIELMRLOTGQGGGLASALDIAETAYQERQGGDPAK
PAVALIHHPIRIQLEARLTETLRATLTLELATWDLAMEPKEBELRGLQMLSTWL
LPLVKNFGADAGFTVNSOAIQILGAGYTEMPLLEOYLARDARVMTIYEGTTGQALDL
LTRLWRGEGKGLDVLKRARESLADYRSDKTKAASVLDAPAESRRMAALENEPET

gene
CDS
GLWQANNVLAAMEAFAMMTVREGVADAQARRTKAA"
/complement (2914..3687)
/gene="AGR_C_2613"
/complement (2514..3687)
/gene="AGR_C_2613"
/note="probable short-chain dehydrogenase PA2554
(imported) - Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_C_2613p"
/protein_id="AAK87207.1"
/db_xref="GI:15156487"
/translation="NNIBGAGALVTGAASGLGAIVARMILARGAAVITFDRNGEAGK
LAAEIGKAVQGDVTDADAQAAIKVAASAKGLRILVNCAGITAGRIIGRGPOPL
GDEQVIRVNLICTFMNMLAAAHMAEREDBESODSRONQVIVTASVAAPFGQICQ
AAYAAKSGCIVSLALPARELARFRIRVNTVAAGITLPLDGLPQEBVESLAGQITPH
PSRLGDPAEFADTVRPLIENDYNGSEVIRLDGAIKMQPR"

gene
CDS
/gene="AGR_C_2614"
/complement (3684..5459)
/gene="AGR_C_2614"
/note="(A0238746) feruloyl-CoA synthetase"
/codon_start=1
/transl_table=11
/product="AGR_C_2614p"
/protein_id="AAK87208.1"
/db_xref="GI:15156488"
/translation="MWRNDPLGPYPQKLNERRLLHWCRAFPERTMADROGPEPMRV
SYAELDKIRRIQFLDHDLSVERPLVLSNSIBHALMVLAAQHVGLASAAITAY
ATSAIDLKADIIGQITPGWFAEDATPPRALGEVFDGTPVGLRNLIPEDSNTFH
FETLLETPBEAVDRAFDVAGPDVAKFLFTSGTGS PKAVIQTORMLGNOMILD
YGVREPEPVVDMAEPMNHTAAGNKVNLVNGCYVYIDRGPSDAICGQTDNLND
ISPITYFNVPAGHEMLVQAMKRDALCRSFRDLKMLMAGAMAGHTDALTELSMA
TVGHAVMGAGLSTETAPSLCTEPQDPKPGNIGIPAGCVTKLVPPDERVELRKG
PNITPGYWRNGELTAAAFDEBGFYRGDVYKFAVDPPRGFYFDGRMMEFNLQGT
WVAVGPRLAQLVMFAGLIRDAVITGENAEALGALVPPFIPALREHVRGSHLSDSEI
IRHPSVAQIVAKLSHOKOASGSARVNRILVMEDALRPEKEGVTDKGS INQRAVIL
HREKIVESLYGDTPQVITVGRVAA"

gene
CDS
/gene="AGR_C_2616"
/complement (5610..5458)
/gene="AGR_C_2616"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_2616p"
/protein_id="AAK87209.1"
/db_xref="GI:15156489"
/translation="MISPPGRFSHATSGDSFTLLIVTSSQTSFTRTVRIIFCQKNDY
KTIISL"

gene
CDS
5664..6461
/gene="AGR_C_2617"
5664..6461
/gene="AGR_C_2617"
/note="enoyl-CoA hydratase (fad-1)"
/codon_start=1
/transl_table=11
/product="AGR_C_2617p"
/protein_id="AAK87210.1"
/db_xref="GI:15156490"
/translation="MSEKYTEEFVSVELRGDIALVGLANRPOKRNATSDRFVEALAA
AAAEENARAAYIFEGHDFRCAGLDGELHHTKPIBEVRSRRKHAFVSOIBHTITPV
SAHAGVAVGGGJELASTHRLRVADDESAFPALPEGGQVIGGGGSVRIARLIGVAKMS
DMULTGRVASSAEARWNLYOVAVPRGELAAQELATTIAASNAIPSNVAIVNALPRI
QMAKEDGILFVSEFTSSPTAASPEAEERLRATLEKRTATLKADDTTA"

gene
CDS
/gene="AGR_C_2619"
/complement (6511..7506)
/gene="AGR_C_2619"
/note="(A0101317) MetP protein"
/codon_start=1
/transl_table=11
/product="AGR_C_2619p"

```

/ protein_id="AAK87211.1"
/db_xref="GI:15156491"
/translation="MVFRLIEVAVSCIHNOVSLREKAMGMSFWMDNSGADGVYRS
SVILGMSIEVMPRTAKIDIPNLIIDGRTVYIAHIDGPIADWSTARIADDEGF
VNPILPARGISGNKELEMDLKRQYEGATYQALLGGGKOPAGSPSSGIDMEITLF
DKMGFTHLIAGHPENRDIIDSGSSRGDOLIMKODASRTDALATTTPAFPAK
VTEWAKRIGMNGVTLIVHGIAGPYKTLIKFALICGVPSLKLQKRALDLSLF
VVEPTPIAEIIDLKANPDSLIEHVHIFPLGIRPSAEWMKQASTGTLANVA"
gene
7739 .8473
/ gene="AGR_C_2620"
7739 .8473
/ gene="AGR_C_2620"
/ note="probable transcription regulator PA4906 {imported}"
/ pseudomonas aeruginosa (strain PAO1)"
/ codon_start=1
/ translation="MROVDAATHGKRAVIELREKILSGELPGMRLFEVSTAEIDIS
RPPVEBALSLTEBGLNRLPGGQFVVRFGFADVVDAIEVHGMEGTARILAEERG
SKVALEIDATVOQLDLCFGRVDVDFGVYALNRI FHHOLAALCGSEMRREVERA
SSLFAPSAPFLPKANIGAPRRSLRAGQEHKAIYAAIVAREGAAAEVAHESHTA
RNLEWMTREARBELIAQVPLALISD"
6623 .10044
/ gene="AGR_C_2622"
6623 .10044
/ gene="AGR_C_2622"
/ note="PROBABLE AMINOMETHYLTRANSFERASE (GLYCINE CLEAVAGE
SYSTEM T PROTEIN)"
/ codon_start=1
/ translation="M1556493"
/ protein_id="AAK87213.1"
/db_xref="GI:15156493"
/ translation="MTMARSILTBVUKOYDNPVEMRNKIGVYVYVAVGESMRS
EQARREGAVLPDOSHMDLVEGSDAKPLAHTGINSFNDLRAHPVPYTPDG
HVIQDHIIFRERQDKFLIVGRAPTSWMLFCAYGKMNRIKHDPSPRPERGLVR
HYRVOIQPDAKPIFEKNNGGPIPDIKFHVDMINIGSKYQALHSGMSGALBWL
GYEDKSYLLSCILEAKQAGVNLVRCGRAYSTNLSJESGIMPSPGPGYTGEMKLD
YDMLGADSYEATGAIKGSFVSENIEDYVNPBELGIDYDIKMKDDPIGKSLKEMK
GKTHRKKTPEMNAEDVIVIASAPRGEDHKMIDIPRANVASTGADILKGGKHYG
MFMPTGYSTNERCIIISLGIYVNDIQIGDVLTLVWGPDPGSAKSTIERKHQEKIRVR
SETPYAAERNDVADSWRTKAS"
ORIGIN
Query Match 22.6%; Score 399.4; DB 1; Length 10173;
Best Local Similarity 52.8%; Pred. No.5.4e-85;
Matches 909; Conservative 0; Mismatches 806; Indels 6; Gaps 2;
QY 12 CGAGGCGCTTCTCCCTTCCCGGTCGAATTCTTGAAGCTCTCGACATTTGGGCTAAGAC 71
Db 5445 CGATCGCTTTGGCCCTTATCCGCAAAATTGAATGAGCGCTCTTCACTGCGTGTGTT 5386
QY 72 CCGTCCAGAACAACTGCGTTCCTCCAGGCGCGCAAAATGGGAAATGCGCTGTATCAG 131
Db 5385 GGCACCGGAGCGCAATGATGCGCCGACGTCAGGCTCCGACCAATGCGCGCGGTTCAG 5326
QY 132 CTACGCGGAATGTTCCACACGTCGCGGCATCGACAGAGCTTCTCTTACGGAAGT 191
Db 5325 CTACGCGGAGCTCTGACAAAGTACGGGATACGGCCAGTCTCTCTGATATATACCT 5266
QY 192 ATCGGAGAGCGCTCGCTGCTTATCGTCTCTGGAATGACCTTGAAATCATCTTCACTG 251
Db 5265 TTCTGTCGAGAGGCGCTGCTGCTTCTTCTGAAAATTCATGACATGACGCTGATAGGT 5206
QY 252 ATTTGGGGCTATGATGCGGCAATTCCTATTCGCCGGTGTCTCTGCTTATTCATGCT 311
Db 5205 GCTTGGGCGCCAGCATGTGCGTATCGCTCGCTGCAATCAGCCGCGCTATGCTC---CAC 5149
QY 312 GTTCGCAAGATTTGGCGAGCTGCGTCAATCGTAGGCTTTCTGCAACCGGGAATGGGCTT 371
Db 5148 GTTCGCGCATCTCACAAACTGCGCGCAATCGCGGCGAGATCACCCCGGAGATGGTATT 5089

```

```

QY 372 TGCTGCCATGAGACACTTTTCCAGCGCGCAATTGAGACCAATTCTGCGCGACGCTGCC 431
Db 5088 TGCCGAGATGAGAGCGCTTTTCGCGGCGCGCTCGCGAGGTTTTCAGACGAGTACACC 5029
QY 432 CGCAATCTTCATCTCGAGGGAAATTGGCCGGGGGGCC---ACGGTAGTTTTCAGACGCT 488
Db 5028 GCTGTGCGGCTCCCGCAACTCGCGGAAGATGTTTCCAAATGAGTTTCATTTTGAAACCTT 4969
QY 489 GCTGAGACAGCTGTGGGATTTAGGACAGATATAGCTTTGGCGCAACTGCGCCGATAC 548
Db 4968 GCTTGAAAGAGAGCGGACCGGAAGCGGTGATGTGCTTTTCAATGCGGTGCGCGCGGATAC 4909
QY 549 GATTCGCAAGTTCTTGTTCATTTCTGCTCTTACCAACTGCTTAAGCGGTGCGCATAC 608
Db 4908 CGTTCGCAAAATTCCTTTTCACTGACGTCAGGACGACGATGCTTGCCTCAAGCGGTATCCAGAC 4849
QY 609 TCAGGCAATGCTCTGCGCCAAATCAGAGATGCTTCTGCAAACTTTCCCGGTTTGTGTA 668
Db 4848 GCMAAGCATGTTGTGCTCAATCAGGAATGATGCGCATTTGTATGCTACTTTCGCA 4789
QY 669 AGAGCGCGGTCGTGTGGACTGTGTGCGTGGAACACACCTTTCGCGGCGGACACAA 728
Db 4788 GAGCGCGCGGTCTGTGATTTGGCGGCAATGAAACACACGCGCGGGAACAAGGT 4729
QY 729 CATCGGATCTGTGTGTAACAAGCGCGGCAAGTACTTGAAGACGCTTAAACCAACCGC 788
Db 4728 GTTCAACCTGCTCTTATATGCGGCTACTATTATATGATGCGGCGCAAGCCGACCC 4669
QY 789 CCAAGGCTTTCGCGAGAGCGCTTCGACATTTGAGCGCAATCTCCACACGCTGACTCAC 848
Db 4668 GCGCAGATGCGCGACACCTGACATCTCGGACATATGCGCCACTGTATTTCAA 4609
QY 849 TGTGCCAAGAGCTGGAAGAAATTAAGTGGTCCCTTGAAGGACAGATACCTTCGCGCA 908
Db 4608 CGTTCCCGTGCACAGAAATCTGTCCAGCAATGCGCAAGACGAGCGGTGTGCGG 4549
QY 909 AGCTTTCTTCTGCTGATAGCTGTTCTTCTTCCGCGCGCTGGTGTCCAGAGGAT 968
Db 4548 CAGCTTTTTCGCGACTTAAATGCTGATGATGTCGCTGCGGCTGCGGATGCGGACAGC 4489
QY 969 CTGGGATCTTTGACACCGGCTGCTGMAAGACATGCTGTGAGCGCATTCGATGATGAC 1028
Db 4488 ATGGAGTGCCTGACGAGACTTTCATGGAACAGCTGCGCATGCGGTCTCATGCGTGC 4429
QY 1029 GGTCTGCGCATGAGAGACATGCTCTCTTCTGCACTTTTACACCGGACCGCTGCAAT 1088
Db 4428 CGGCTCGGCTGACGAGAACGGAACCTTCTGCGTGTCTGACACGAACCGAGGACAA 4369
QY 1089 GGTGTGTTACATTTGGCTGCAAGCGCTGCTGCGAGTCAAGCTTCTGCTGATGG 1148
Db 4368 GCCCGCAATATCGGATCTCCGCGCGAGGCGTGAAGTGAATCTGTCCTTCATGAG 4309
QY 1149 GAAATTTGAAGGCGTTCATAGTTCGCGACGTCATGACGCGCTACTGCGGTCTCTGA 1208
Db 4308 CCGGTACGAATTAACGCTGAAGGCCCAACACACCGCGGTATCTGCGATAGCGCA 4249
QY 1209 ACAAATGCGCAAGGTTTCGACGAGGAAGCTATTACTGCTCGGTGATGCCATCAAT 1268
Db 4248 ATTTAGCGGCGGACGCTTTCGACGAGGAAGGTTCTATGCAATCGGGAATACGTCGAAAT 4189
QY 1269 GGCAGATCTGCGCATCTCAAGAAAGTGTATGTTTGAAGGTCGATTTGGAAGACTT 1328
Db 4188 TCCCGTCCGCGACGATCCCGCGCGGCTTTTATTTTCATGTCGCGCATGCGGAAACTT 4129
QY 1329 CAAGCTGCTCAAGGGATTTTGTCAAGTGTGGGCAATTTGCGCAAGCGGCGTTCGGA 1388
Db 4128 CAAGCTGCAACCGGACCTGAGTGGCGGTCCGCGCTTTTGGCGGCAATTAETCAACAT 4069
QY 1389 AGCGGCTTTTACGTCCTTGAAGTATGAGTGTCTGCTCTGTAATGCTGTTGATTT 1448
Db 4068 GTTCCGCGGCTGATTCGTGATGCGCTCATTAACGCGGCAAAACCGGCGCAACTGCGGCGC 4009

```

```
Qy 1449 GTCGATGTTCCGCGTCTTCCGACTGCGCGCTTGTCCGGGCTAGAAAGAGCGTC 1508
    |||||
Db 4008 TCTGTCGTGTCCTTCAATTCATACCGGCTACGTGAGCTTGTCCGGTTCGCAACATCTCTC 3949
    |||||
Qy 1509 GGAAGCCGAGAGTCTTCCAGTAGAGCCGATTCCGGGCTGTGCTGACTGGCTCAACG 1568
    |||||
Db 3948 CGATGCGGAGATCATCCGACCCGAGTGTCAAGGCGCAGATTGTGCGAAGCTTTCGGC 3889
    |||||
Qy 1569 ACTCAATCGAAGAACAACTGGCAATGCCAATGCCATCATGTGGGTAGGGCTCCTCGATAC 1628
    |||||
Db 3888 GCATCAAAAAACAGCGGCGGTTCCGCAAGCCGCGTCTATGCGCATTCGTGATGGAAGA 3829
    |||||
Qy 1629 GCGCGGTGATTAAGGCGAGTCACTGACAAAGGCTCGATCAACGAGCGGCTGT 1688
    |||||
Db 3828 CGCGCTGCGCTTCGAGAAAGGCGAGTTTACGACAAAGGATCGATCAACCAAGCGCGCGT 3769
    |||||
Qy 1689 TTGCAATGCGGCTGCGGCAAGTTGATGCGCTGTATCGTG 1729
    |||||
Db 3768 GCTTGTGCATCGCAAGGAGCTCGTAGAGTCACTTTACGGCG 3728
    |||||
```

Search completed: February 7, 2005, 20:09:44
Job time : 7371 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2005, 23:26:45 ; Search time 44 Seconds
(without alignments)
887.757 Million cell updates/sec

Title: US-09-750-986D-30

Sequence: 1 MRSLEALLPPGRILRLERLH.....KVDAIYRGDSMLRDEATL 589

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3082	100.0	589	US-08-976-063E-30	Sequence 30, App1
2	1652.5	53.6	633	US-09-328-352-6519	Sequence 6519, App1
3	300	9.7	457	US-08-311-731A-52	Sequence 52, App1
4	298	9.7	488	US-08-311-731A-283	Sequence 283, App1
5	266	8.6	614	US-09-543-681A-7066	Sequence 7066, App1
6	263.5	8.5	544	US-09-396-154-23	Sequence 23, App1
7	263	8.5	634	US-09-418-963-3	Sequence 3, App1
8	262.5	8.5	544	US-09-396-154-24	Sequence 24, App1
9	262.5	8.5	544	US-09-396-154-45	Sequence 45, App1
10	257.5	8.4	544	US-09-396-154-44	Sequence 44, App1
11	255.5	8.3	544	US-09-396-154-21	Sequence 21, App1
12	254.5	8.3	544	US-09-396-154-22	Sequence 22, App1
13	252.5	8.2	544	US-09-396-154-19	Sequence 19, App1
14	251.5	8.2	544	US-09-396-154-20	Sequence 20, App1
15	249.5	8.1	550	US-09-602-628-2	Sequence 2, App1
16	247.5	8.0	550	US-09-602-628-4	Sequence 4, App1
17	246.5	8.0	550	US-08-354-240A-4	Sequence 4, App1
18	246.5	8.0	550	US-08-354-240A-6	Sequence 6, App1
19	246.5	8.0	550	US-08-718-425-5	Sequence 5, App1
20	246.5	8.0	550	US-09-602-628-6	Sequence 6, App1
21	246.5	8.0	550	US-09-602-628-8	Sequence 8, App1
22	246.5	8.0	550	US-09-577-424-2	Sequence 2, App1
23	244.5	7.9	616	US-10-140-002-206	Sequence 206, App1
24	243.5	7.9	815	US-08-122-520C-9	Sequence 9, App1
25	242.5	7.9	815	US-08-875-277A-4	Sequence 4, App1
26	241.5	7.8	550	US-08-718-425-2	Sequence 2, App1
27	240.5	7.8	549	US-08-354-240A-2	Sequence 2, App1

28	240.5	7.8	550	3	US-08-867-352-23	Sequence 23, App1
29	240.5	7.8	550	3	US-09-380-0618-21	Sequence 21, App1
30	240.5	7.8	550	4	US-09-602-628-12	Sequence 12, App1
31	240.5	7.8	550	4	US-09-396-154-31	Sequence 31, App1
32	240.5	7.8	1242	4	US-09-488-270A-2	Sequence 2, App1
33	237.5	7.7	550	3	US-08-875-277A-2	Sequence 2, App1
34	237.5	7.7	550	3	US-08-487-183A-10	Sequence 10, App1
35	237	7.7	739	4	US-10-140-002-478	Sequence 478, App1
36	236.5	7.7	512	4	US-09-602-787A-86	Sequence 86, App1
37	236.5	7.7	544	4	US-09-396-154-14	Sequence 14, App1
38	236.5	7.7	568	4	US-09-602-787A-30	Sequence 30, App1
39	234.5	7.6	544	4	US-09-396-154-17	Sequence 17, App1
40	234	7.6	538	4	US-09-489-039A-13518	Sequence 13518, App1
41	233	7.6	582	2	US-08-403-852D-20	Sequence 20, App1
42	233	7.6	582	2	US-08-510-646B-21	Sequence 21, App1
43	233	7.6	582	3	US-09-231-818-20	Sequence 20, App1
44	233	7.6	582	3	US-09-635-359B-20	Sequence 20, App1
45	232.5	7.5	550	3	US-09-380-0618-6	Sequence 6, App1

ALIGNMENTS

```

RESULT 1
US-08-976-063E-30
; Sequence 30, Application US/08976063E
; Patent No. 6524831
; GENERAL INFORMATION:
; APPLICANT: Steinduchel, Alexander
; APPLICANT: Pfeifer, Horst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
; TITLE OF INVENTION: ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; FILE REFERENCE: VANILLIC ACID AND THEIR USE
; CURRENT APPLICATION NUMBER: US/08/976, 063E
; CURRENT FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 30
; LENGTH: 589
; TYPE: PRT
; ORGANISM: not required under old rule
US-08-976-063E-30

Query Match      100.0%; Score 3082; DB 4; Length 589;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSLEALLPPGRILRLERLHMAKTRPEQTCVAAARANGERRRISYAMFHNRAIOSL 60
DB 1 MRSLEALLPPGRILRLERLHMAKTRPEQTCVAAARANGERRRISYAMFHNRAIOSL 60
QY 61 PYGSAERPLIYSGNDLEHLQAFGMVAGIYCPVSPAYSLSDLAUKRHVGLQP 120
DB 61 PYGSAERPLIYSGNDLEHLQAFGMVAGIYCPVSPAYSLSDLAUKRHVGLQP 120
QY 121 GLVFAADAAFFORAIETILPDDVPAITRREGELAGRITVSFDSLEPGIEADNAFAATG 180
DB 121 GLVFAADAAFFORAIETILPDDVPAITRREGELAGRITVSFDSLEPGIEADNAFAATG 180
QY 181 PDITAKLFTSGSTGLPKAVPTTORMICANQOMLQTFPVGGEPPVLVDMLPNNHTFGG 240
DB 181 PDITAKLFTSGSTGLPKAVPTTORMICANQOMLQTFPVGGEPPVLVDMLPNNHTFGG 240
QY 241 SHNIGIYVNGGYLDDGKPTAGFPAETLRNISESPYLYITVPGKMBELVGLERDST 300
DB 241 SHNIGIYVNGGYLDDGKPTAGFPAETLRNISESPYLYITVPGKMBELVGLERDST 300
QY 301 LRERFFARMKLFPAAGLSQGIWDRLDRAVBOHCGERIRMMAGLMTETAPSCFTTGP 360
DB 301 LRERFFARMKLFPAAGLSQGIWDRLDRAVBOHCGERIRMMAGLMTETAPSCFTTGP 360

```

Db 301 LRRERFARMKLFEEAAGLSQGIWDRLDRAEOHCGERIRMMAGLGMTETAPSCFTTGP 360
Qy 361 LSNAGYIGLPAPGCEVTLVPDGLRPGPHMGVSGWMPAPBONAOAFDEEGYCSGDA 420
Db 361 LSNAGYIGLPAPGCEVTLVPDGLRPGPHMGVSGWMPAPBONAOAFDEEGYCSGDA 420
Qy 421 IKLADPADPKGLMPDGRIRIADPFKLSGCVSVSGVPLTRAVLEGGSYLVDDVVAAPDREC 480
Db 421 IKLADPADPKGLMPDGRIRIADPFKLSGCVSVSGVPLTRAVLEGGSYLVDDVVAAPDREC 480
Qy 481 LGLVFPRLDLCALSGLEKESAPAEVLASEPVRAMPADWLKTLNRATNGASIRIMVGL 540
Db 481 LGLVFPRLDLCALSGLEKESAPAEVLASEPVRAMPADWLKTLNRATNGASIRIMVGL 540
Qy 541 LDTPEPIDKGEVTDKGSINORAVLQWRSKADVALYRGEDQSMLEDEATL 589
Db 541 LDTPEPIDKGEVTDKGSINORAVLQWRSKADVALYRGEDQSMLEDEATL 589

RESULT 2

US-09-328-352-6519
; Sequence 6519, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6519
; LENGTH: 633
; TYPE: PR1
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6519

Query Match 53.6%; Score 1652.5; DB 4; Length 633;
Best Local Similarity 53.4%; Pred. No. 6.5e-168;
Matches 310; Conservative 100; Mismatches 168; Indels 3; Gaps 3;

Qy 5 EALLPFGRLERLEHMAKTRPEOTCVAAAPANGEMRISYAEHFNVRALIAQSLPYGL 64
Db 45 EOLKPYQKLTDRILHFAQTKRPHIFAKRNAQOEHWKLSYAEVLORAMHIAQALHARNL 104
Qy 65 SAREPLIVSGNDLEHQLAFGAMVAGIPYCPVSPAYSLSODLAKLRHIVGLQPLGVF 124
Db 105 SGERPLVITLSEGNLEHLLTLMAMLAGVPSPAYSLSODPGLKHYFEVLTGPMVY 164
Qy 125 AADAAPFORAIEFTLPDDVPAIFTRGELAGRRTVSPDLSLEOPGSLADNAFAATGPDIT 184
Db 165 ASDGQAFKAIAQCTPIDIEVNTKGIYDQICTSFSLDTPVS-NVGEFYQTLDBNOI 223
Qy 185 AKELFTSGSTKLRKAVATTQRMICANQOMLLQTFPVGSEPPVLVMDLPMNHTFGGSHN 244
Db 224 AKELFTSGSTKLRKAVATTQRMICANQOMLLQTFPEEETPPVLLDLSMHTFGGSHN 283
Qy 245 GIVLYNGGTYYLDGKPTAOGFAETLNLSEISPTALVTPKGEELVGLERDSTLRER 304
Db 284 GIALYNGGTYYLDGKRVAKRFDETINLKEISTVYLVANPKGHEELTEALEKDEELRER 343
Qy 305 FFAARMKLFEEAAGLSQGIWDRLDRAEOHCGERIRMMAGLGMTETAPSCFTTGPLSMA 364
Db 344 FFAARMKLFEEAAGLSQGIWDRLDRAEOHCGERIRMMAGLGMTETAPSCFTTGPLSMA 403
Qy 365 GYIGLPAPGCEVTLVPDGLRPGPHMGVSGWMPAPBONAOAFDEEGYCSGDAIKL 423
Db 404 GYIGLPAPGCEVTLVPDGLRPGPHMGVSGWMPAPBONAOAFDEEGYCSGDAIKL 463
Qy 424 ADADADPKGLMPDGRIRIADPFKLSGCVSVSGVPLTRAVLEGGSYLVDDVVAAPDRECGL 483
Db 464 VDDVVAAPDRECGLMPDGRIRIADPFKLSGCVSVSGVPLTRAVLEGGSYLVDDVVAAPDRECGL 523

Qy 484 LVPRRLDLCALSGLEKESAPAEVLASEPVRAMPADWLKTLNRATNGASIRIMVGL 542
Db 524 LVPRRLDLCALSGLEKESAPAEVLASEPVRAMPADWLKTLNRATNGASIRIMVGL 583
Qy 543 TPEPIDKGEVTDKGSINORAVLQWRSKADVALYRGEDQSMLEDEATL 589
Db 584 TPEPIDKGEVTDKGSINORAVLQWRSKADVALYRGEDQSMLEDEATL 624

RESULT 3

US-08-311-731A-52
; Sequence 52, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 457 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM LEPRAE
US-08-311-731A-52

Query Match 9.7%; Score 300; DB 4; Length 457;
Best Local Similarity 26.2%; Pred. No. 5.5e-23;
Matches 143; Conservative 67; Mismatches 210; Indels 126; Gaps 27;

Qy 44 SVAEHNNRAIQSLPYGLSAERLLTVSGNDLEHQLAFGAMVAGIPYCPVSPAYSL 103
Db 13 SVAEHNNRAIQSLPYGLSAERLLTVSGNDLEHQLAFGAMVAGIPYCPVSPAYSL 157
Qy 104 LSODLAKLRHIV-----GLQPLVPAADAPFORAIEFTLPDDVPAIFTRGELAGRRT 157
Db 58 VER-----RHMLDSGAQAMLGGR--SADSA-----PDGLPHI-----P 90
Qy 158 VSFDSLLEOPGCEVTLVPDGLRPGPHMGVSGWMPAPBONAOAFDEEGYCSGDAIKL 217
Db 91 VQDLA-----RSWNRYPESPDPATAMVITYSGTGPCKGLVLSRRALAVDLALQA 142
Qy 218 FPFVGEPPVLYDMLPMNHTFG-----GSHNIGIVLYNGGTYYLDGKPTAOGFAETLR 271
Db 143 WQNTAAD--VLVHGLPLFHVHGLVGLGLSLRI-----GNRFVHTGKPTTAIVQA-- 191

QY 272 NLSEISPTAYLVTPKGEELVGALEBDSTLREFFARMKLPFPAAGLSGIMDRDLRVA 331
 DB 192 -CSBAGSLYFGVPTVWSRLVA---DEAV-ARALRPARLLVSGASLPVPVFDRLAHLT 245
 QY 332 EOHGGERIRRMAGLGMTETAPSCFTTGPLSMAGYIGLPAPGCEVKL-----VPVDGK 384
 DB 246 GHRPIER-----YGSTESLITLSTLADGERBRAGWGLPLAGVOTRLVDESGGVPPYDGE 299
 QY 385 LEGRF--HGPHWNGSYWRAPFQNAQAFDEEGYCCSGDAIKLADPADPQKGLMFDGRIAD 442
 DB 300 TVRQLVRSPTMGFYLNREPEATAEAFDEDEGWYRTGD--VAVVDSGGMHR--IVGRESVD 355
 QY 443 FKLSGVFVSVGLRTFAVLEGSSYVLDV--VVAAPDEECGLVFPRLDORALSGLGE 501
 DB 356 LILKGGRIGAGSI--EMALLGHPDVREAVAVGLPD--EDLGQRIVAFV-----VGAE 404
 QY 502 ASDAEVLASBPVAMFADMLKRLNREATGNASRIWVGLDTPPSIDKEVTDKGSINOR 561
 DB 405 ALDADELIN-----YVAQQLSIHKRPRE-----VRPDLALPA--TRWERSKSSCSSR 450
 QY 562 AVLQWR 567
 DB 451 A--DMR 454

RESULT 4
 US-08-311-731A-283
 ; Sequence 283, Application US/08311731A
 ; Patent No. 6583266
 ; GENERAL INFORMATION:
 ; APPLICANT: SMITH, DOUGLAS
 ; APPLICANT: MAO, JEN-1
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
 ; NUMBER OF SEQUENCES: 411
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
 ; STREET: 600 ATLANTIC AVENUE
 ; CITY: BOSTON
 ; STATE: MASSACHUSETTS
 ; COUNTRY: USA
 ; ZIP: 02210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/311,731A
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: GATES, EDWARD R.
 ; REGISTRATION NUMBER: 31,616
 ; REFERENCE/DOCKET NUMBER: C0044/7125
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/720-3500
 ; TELEFAX: 617/720-2441
 ; INFORMATION FOR SEQ ID NO: 283:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 488 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Mycobacterium leprae
 ; US-08-311-731A-283
 Query Match 9.7%; Score 298; DB 4; Length 488;
 Best Local Similarity 26.4%; Pred. No. 1e-22;

Matches 135; Conservative 61; Mismatches 196; Indels 120; Gaps 24;
 QY 44 SYAEMFNHNPALAQSLPFLYLSAERPLIYSGNDLHQLAFAQWAGIYCVSPAYSL 103
 DB 50 SVABRVGAGALVAVLATP--TASTVALAT-----CCLLAGVVVPVADIGV 94
 QY 104 LSODLAKLRHIV-----GLLOGLVFADAPFORALETILPDDVPAIFTRGELARRT 157
 DB 95 VER-----RMLUDSGAQLGRCF--SADSA-----FDGLPHI-----P 127
 QY 158 VSPDSLLEPGGIEADNFAFAPDPTIAKFLFTSGSTKLPKAVPTTQRMICANQMLLOT 217
 DB 128 VQDLA-----RSMNRVPEPSPDPAWAVIYSGTGPCKGVLSRAIAVLDLALAOA 179
 QY 218 FVPGEERPVLYDMLPNHTFG-----GSHNIGIYLVNGCTGYLDDGKRTAGFAETLR 271
 DB 180 WQMTADP--VLVHGLPLFHHGLVLGLLSGLRI-----GNRPVHTGKPTPAVQA-- 228
 QY 272 NLSEISPTAYLVTPKGEELVGALEBDSTLREFFARMKLPFPAAGLSGIMDRDLRVA 331
 DB 229 -CSBAGSLYFGVPTVWSRLVA---DEAV-ARALRPARLLVSGASLPVPVFDRLAHLT 282
 QY 332 EOHGGERIRRMAGLGMTETAPSCFTTGPLSMAGYIGLPAPGCEVKL-----VPVDGK 384
 DB 283 GHRPIER-----YGSTESLITLSTLADGERBRAGWGLPLAGVOTRLVDESGGVPPYDGE 336
 QY 385 LEGRF--HGPHWNGSYWRAPFQNAQAFDEEGYCCSGDAIKLADPADPQKGLMFDGRIAD 442
 DB 337 TVRQLVRSPTMGFYLNREPEATAEAFDEDEGWYRTGD--VAVVDSGGMHR--IVGRESVD 392
 QY 443 FKLSGVFVSVGLRTFAVLEGSSYVLDV--VVAAPDEECGLVFPRLDORALSGLGE 501
 DB 393 LILKGGRIGAGSI--EMALLGHPDVREAVAVGLPD--EDLGQRIVAFV-----VGAE 441
 QY 502 ASDAEVLAS-----BPVAMFADMLK 523
 DB 442 ALDADELINVAQQLSIHKRPREVRFDALPR 473

RESULT 5
 US-09-543-681A-7066
 ; Sequence 7066, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETTON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709, 1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 7066
 ; LENGTH: 614
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-7066
 Query Match 8.6%; Score 266; DB 4; Length 614;
 Best Local Similarity 21.6%; Pred. No. 4e-19;
 Matches 140; Conservative 105; Mismatches 270; Indels 132; Gaps 23;
 QY 14 ILERLEHMAKTRBQTCVAAAPANGEMKRISYVEMFNHNPALAQSLPFLYLSAERPLIY 73
 DB 24 IIRRLQWQFSSSGKTYARQWBAKSB-IAMSWAVAKTRALNALDMGVAVAVQENVGIF 82
 QY 74 SGN-----DLHLQALFAPWAGIYCVSPAYSLSDLKLRIHIVGLQGLVFAA 126
 DB 83 SONSIDMSIADIATLQLR-----AVTVPIYATSSVEQA--AYIINDADIRILIFVG 130
 QY 127 DAAPFQRAIR-----TLPDDVPAIFTRGELARRTVSPDSLLEPGG 169

Db 131 DOKEYDVSESLALLCPQLKHIIIVENSQVNLNPTIPSYHLTNLI SDQS -OYDLSLQO--- 186
Qy 170 IEADNFAATGPPTIAKFLFTSGSTKLPKAVPTTORMLCANOQMLQTFPVGSEBPVLV 229
Db 187 -RIDECL---DDLFLITSTGTGEPKVMLDYTLAS--QYLDQSLSDKQVSL 239
Qy 230 DMLPMNHTFGSHNIGIVLNGSTYLLDDKPTAOGFAETLRNLSEISPTAVLTVPKGWE 289
Db 240 CFLPLSHVFERRAWSFYVMHTGALINVYLTDTHAVREANA-----AVKPTVMCAVPRFYE 292
Qy 290 ELVGALEBDSLTAEERFAARKLFFFA-----AAGLSOGI---W-----DRDR 329
Db 293 KYVSALIOEKVSQAPRL--RQMFMRMLKQGEKORQAHLARSLGVSQWCFYRPAKKVLQ 350
Qy 330 VAEQHCGERIRMA-----GLQMTETAPSCFTTTPBLSMAGY 366
Db 351 PLRQIIGGRVRFPPAGARLDVAIGFPLAAGINIKKYGTMTCTCATVSCWENHYLGSS 410
Qy 367 IGLPAPCEYKLVVPDQKLEGRFPHVMSGYWRAPQNAQAFDEBGYCGDAIKLADP 426
Db 411 IGTPLGVEYR---IGBENEIOVRGSIVMKGYFNKPETVAALFTEDQMLRTGAGAL--- 464
Qy 427 ADQKGMFPGRIAEADPKLSSGVFVSGLRTAVLEGGSYLDVVVAADRECLGLV 466
Db 465 -DEQMLYITERLKDLMKTSNGKYIA--PQMEGTIGQDFRIFHIAVIAPTKRFVSALIV 521
Qy 487 PRLLDC-----RALSGKEASDAEVLASEPVRAMPADMLKRLREATGNASRIMV 538
Db 522 P----CFDLEERARALN--LKYHDRHELRHKIKELFERLREMRN--FASFHQVKR 574
Qy 539 GLLDTPPSIDKGEVTDKGSINQRAVLQWRSAKVADALVGGEDQMLRD 585
Db 575 TLAEFTWESGELT-----PTLKRRIKISERYRNEIEQWQE 613

RESULT 6
US-09-396-154-23
; Sequence 23, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; FILE REFERENCE: 341.012US1
; CURRENT APPLICATION NUMBER: US/09/396,154
; EARLIER FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: US 09/156,946
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant luciferase
; NAME/KEY: UNSURE
; LOCATION: (354)...(355)
; OTHER INFORMATION: Unknown amino acids
US-09-396-154-23

Query Match 8.5%; Score 263.5; DB 4; Length 544;
Best Local Similarity 22.5%; Pred. No. 6,1e-19;
Matches 130; Conservative 86; Mismatches 226; Indels 137; Gaps 24;
Qy 30 CVAARANGEMKRISIAEMHNVRALIAQSLPYGLSARERPLLVSGDLEHQLARQAMV 89

Db 38 CIALTNATNR-ENVLYEEFLKLSCLRAESFKKYGLKQNDTIAVCSENGLOFLPLVIASLY 96
Qy 90 AGIPYCPSPAYSLSLQDLAKLHIVGLQPLGVFAPADAAPFORALETILLDDVPAIFTR 149
Db 97 LGIIVAPVNDKY--IREELI---HSLGIYKPRIFPESKRT--PQKVLN-----VK 139
Qy 150 GELAGRTVSPSLLEBQPGIEADNAPAAATG-----BDTAKFLFTSG 192
Db 140 SKLKSVEITIIILDNLEDGLGYCQANFISQNDINDVKKKFKPYSFRDDQVALIMFSSG 199
Qy 193 STGLPKAVPTTQRMLCANOQMLLQTFPVG---EEPPVLVDMLPMNHTFGSHNIG--- 245
Db 200 TGLPFGVNLTHKNIVA--RPSLAKDPTFGNAINPTTALITVLVFRHGGMMTTLLGYFTC 257
Qy 246 ---IVLYNGSTYLLDDGKFTPAQGAET--LRNLSEISPTAVLTVPPGMEELVGALEBRDT 300
Db 258 GRRVULMH-----TFEEKLPLQSIQDVKVESTLLV---TLMAFLAKSL 299
Qy 301 LRERFPAARKLFPFAAAGLSQGIWDRLDRAVEQHCGERIR-----MMAGLQMTETAPB 353
Db 300 VEKYDLSHLKEIASGGAPLSKEI-----GEMVKRPFKLNFVRQGYGLTETTS 347
Qy 354 CTFTTPPLSMAGYIGLPAPCEYKLV-PYDQKLEG-----RFHGHVMSGYWRAPQON 405
Db 348 VLTTPKXXARPGSTGKIVFHAQKVDPPTGKILGNEBEGELYFKGPMIMKKGYNNEAT 407
Qy 406 AQAFDEGYCGSDAIKLADPADPQKLMFDEGRIADFKLSGVFVSGLRTAVLEGG 465
Db 408 KAIIDNDGMLRSGDI-----AYYD-----NDGHFIYVDRLSLIXKG- 445
Qy 466 SYVLADVVAAPRECLGLLVPRLLDCRALSGKKEASDAEVLASEPVRAMPADMLKRLN 525
Db 446 ---YQVAPAEIEGI-LIQHPYIVDA-GVTGIPDEAAGELPAAQVAVVQTG---KYLN 493
Qy 526 RE-----ATGNASRIMV-----VGLLDTPPSIDKGEVTDK 555
Db 494 EQIVQDVSSQVSTAKMLRGVAFLEDEIPKSGTGKIDRK 532

RESULT 7
US-09-418-963-3
; Sequence 3, Application US/09418963
; Patent No. 6664039
; GENERAL INFORMATION:
; APPLICANT: Min, Kyung-Tai
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING
; FILE REFERENCE: 06618-367001
; CURRENT APPLICATION NUMBER: US/09/418,963
; EARLIER FILING DATE: 1997-10-14
; PRIOR APPLICATION NUMBER: US 60/104,298
; PRIOR FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 634
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-963-3

Query Match 8.5%; Score 263; DB 4; Length 634;
Best Local Similarity 21.4%; Pred. No. 8,9e-19;
Matches 135; Conservative 93; Mismatches 282; Indels 120; Gaps 20;
Qy 39 EWRRIYAEHFHNVRALIAQSLPYGLSARERPLLVSGDLEHQLAFGAMYAG---IPY 94
Db 39 KWEHISYQYUULARBAKGPLKLGKQASVALIFQNSPEWFFSIVGVFAGIVTGIX 98
Qy 95 CPVSP-AYSLSODLAKLHIVGLQPLGVFAPADAAPFORALFTII----- 139
Db 99 TTSSPACOTIAYDC-----ANVIWDT---OKOLEKILIKQPLHLAAVVI 144

```

Oy 140 -----PDDVPAFTFGELGRRTVSDSLGEOGGIEADNAPFAATPDITAKLFFNGST 194
Db 145 YKEPPNKGANVYTBEE-----FMEIGNEVPEEALDAI1DITQPNOCVLYVTSGT 196
Oy 195 KLPKAVPTTQRMLCANQOMLQTFPVYEGEE--PVLVDM,PMNHTFGSGSHNIGIVLYNGG 252
Db 197 GNPFKVMLSQDNI1TWTARYGSOAGDIRPRAVEOEVVVSYPLSHIAQIYDLMGTG1QWGA 256
Oy 253 TYULDDGKPTAGFAETLNLSEISPTAYLTVYPKGMEBVLGALERDSTLERPFAIMKLF 312
Db 257 QVCFAPBPDALKSLVWTLR---EVEFTSHMGVPRVWEKIMERIQ-EVAASGSTRKKML 312
Oy 313 FFAAAGLSQGI-----WDRLDRAVE--QH- 334
Db 313 WAMSTLEQNLCLPCPSDLEFPTTRLADYLVLAKEVROALGFAKQCNKPYGAPMMAESTOHF 372
Oy 335 -GGERIRMMAGIGMETETAPSCFTTGPPLMAGY-IGLPAKCEVUKVPVDGKLEGR--FH 390
Db 373 FLGLNIRLXYAGIGLETS-GPHMSPFYRYLSSGLVPGCKVLYVNDABEIGELCLW 431
Oy 391 GPHVNSGYWRAEQNAQAFDEEGYCCSDAIKLABDAPQKGMIFDGR1AEDFKLSGVF 450
Db 432 GRTIFMGYLNMDKTCCEAIDECEGMLHTGBAGRL---DADGFLXYIGRLKELIITGAGEN 487
Oy 451 VSVGLRTRRAVLEGGSYVLDDVVVAAPBDECEIGLVF-----PRLD-----C 492
Db 488 VPPVETIEEAVKHE-LPIISNAILIGQKRFSLSLTLKCTLDBDTSBDQTDNLTEQAVFEC 546
Oy 493 RALSGIGKEASDAEVLASEPVRAMPADMLKRLNREATGNASRIWWGGLDPTPSIDKEV 552
Db 547 QRVGSRATTVSEIIEKQDAVYQAIIEGIRRVVMNMAAPRYHIQKWAILERDPSISGGEL 606
Oy 553 TDKGSINGRAVYQWMSAKDALYRGDSQM 582
Db 607 GPTMKLKRULVLEKYGIIISFY--OEQOM 634

```

```

1      RESULT 8
2      ; Sequence 24, Application US/09396154
3      ; Patent No. 6602677
4      ; GENERAL INFORMATION:
5      ; APPLICANT: Wood, Keith V.
6      ; APPLICANT: Hall, Mary P.
7      ; TITLE OF INVENTION: Thermostable Luciferases and methods of
8      ; FILE REFERENCE: 341.012US1
9      ; CURRENT APPLICATION NUMBER: US/09/396,154
10     ; CURRENT FILING DATE: 1999-09-15
11     ; EARLIER APPLICATION NUMBER: US 09/56,946
12     ; EARLIER FILING DATE: 1998-09-18
13     ; EARLIER APPLICATION NUMBER: PCT/US98/19494
14     ; EARLIER FILING DATE: 1998-09-18
15     ; EARLIER APPLICATION NUMBER: US 60/059,379
16     ; EARLIER FILING DATE: 1997-09-19
17     ; NUMBER OF SEQ ID NOS: 93
18     ; SOFTWARE: FastSeq for Windows Version 3.0
19     ; SEQ ID NO 24
20     ; LENGTH: 544
21     ; TYPE: PRT
22     ; ORGANISM: Artificial Sequence
23     ; FEATURE:
24     ; OTHER INFORMATION: Mutant Luciferase
25     ; US-09-396-154-24

```

Query Match 8.5%; Score 262.5; DB 4; length 544;
Best Local Similarity 22.6%; Pred. No. 7.8e-19;
Matches 131; Conservative 85; Mismatches 226; Indels 137; Gaps 24

Cy 30 CVAAPANGEMRIRIYAEMFNHVRVIAQSLPYGISAERPLIVSGNDLHQAFGMAY 89
 | | | | |
Db 38 CIALTNATK--ENVLYEEFLTKSCRCLASSFKRYGGKQNDTJAVCSNGNQLPFIYALY 96
 | | | | |

```

0Y  AGI P C P S P A S S L S O D A K R N H V G L O P C L V F A A D A P O R A E T I L P D V P A I F R 149
Db  97 L G I Y A P N D K Y - I E R E L I - H S L G I V K P R I V C S K N T - P Q K V L N - - - - - V X 139
0Y  150 G E L A G R R V S P D S L E O P G C I E A D N A F A A T G P D T - - - - - I A K F L T G 192
Db  140 S K L S K I E T I I L D N E D L G G Y O C L N F I T S O N S D M L D V K F K P R S F N R D D O V A L I M F S S G 199
0Y  193 S T K L P K A V P T O R M C A N O O M L O T F P V E G - - - E E P R V L V M L P M N H T F G S H N I G - - - 245
Db  200 T T G L P K G M L T H K I V A - R S L A D P R T P G N A I N P T A L I Y I P R H N G M M T T L G Y F T C 257
0Y  246 - - - I V L N G G Y T Y L D D G K P T A O G F A E T - L R N L S E I S P A Y L I T V P K G M E L Y G A L E R D S T 300
Db  258 G F R V L A M H - - - - - T F E E K L F L O S L O D Y K V E S T I L V P - - - T L M A F L A K S A L 299
0Y  301 L R E F F A M K L F F F A A G S I G I W D R L R V A R Q H C G E R I - - - - - M M A G L G M T E T A S 353
Db  300 V E K Y I D S H I K E I A S G A P L S K E I - - - - - G E M V K K R F K L N F V N O G I G L E T T S A 347
0Y  354 C T F T T G P L S M A G Y I G L P A P G C E V K L V - P V D G K L E G - - - - - R F H G P H V M S G Y M R A P E O N 405
Db  348 V L I T P K G D A K P G S T E K I V P F H A V K V V D P T T G K I L P N E R G E L Y F P K G P M T M K G Y N N E A T 407
0Y  406 A O A P E B E Y V C S G D A I K L A D P A D P Q K G I M P O R I A E D F L S S G V P V S V G P L R T R A V L E B G 465
Db  408 K A I I D N G L A S G D T - - - - - A Y V D - - - - - N U G H F Y I V R L S L I K Y K G - 445
0Y  466 S Y L V D V V A A P D R E C L I G L I V F P R L L D C R A L S G I G K E A S D A E V L A S E P V A M F A D M L K R I N 525
Db  446 - - - - - Y O V A P A E I E G I - L I O H P Y I V D A - G V T G I P B E A G E L P A A G V V O T G - - - K Y L N 4933
0Y  526 R E - - - - - A T G N A S R I M W - - - - - V G L D T P S I D K G E V T D K 555
Db  494 E O I V O D Y V A S O V S T A K M L R G C V K F P D E L I P K G S T G K I D R K 532

```

```

, RESULT 9
, US-09-396-154-45
, Sequence 45, Application US/09396154
, Patent No. 6602677
, GENERAL INFORMATION:
, APPLICANT: Wood, Keith V.
, APPLICANT: Hall, Mary P.
, TITLE OF INVENTION: Thermostable luciferases and methods of
, TITLE OF INVENTION: production
, FILE REFERENCE: 341.012US1
, CURRENT APPLICATION NUMBER: US/09/396,154
, CURRENT FILING DATE: 1999-09-15
, EARLIER APPLICATION NUMBER: US 09/156,946
, EARLIER FILING DATE: 1998-09-18
, EARLIER APPLICATION NUMBER: PCT/US98/19494
, EARLIER FILING DATE: 1998-09-18
, EARLIER APPLICATION NUMBER: US 60/059,379
, EARLIER FILING DATE: 1997-09-19
, NUMBER OF SEQ ID NOS: 93
, SOFTWARE: FastSeq for Windows Version 3.0
, SEQ ID NO 45
, LENGTH: 544
, TYPE: PRT
, ORGANISM: Artificial Sequence
, FEATURE:
, OTHER INFORMATION: Mutant luciferase
, US-09-396-154-45

```

```

Query Match 8.5%; Score 262.5; DB 4; Length 544;
Best Local Similarity 22.5%; Pred. No. 7, 68-19;
Matches 130; Conservative 86; Mismatches 226; Indels 137; Gaps 24;

QY 30 CVAARAANGERRRISTYAEFMFNVAIAQSLPIYGISAEERPLIYSGNDLHQLAFGM 89
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 38 CIALTNATK-ENVLYEEFLKSLRLESPFKYGIKONDTIAVCSNLSQFFLPVIL 96
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```
QY 90 AGIPYCPSPAYSLSLQDLAKLHIVGLQPGLVFAADAAPFORALETLLPDDVPAIFTR 149
|||
|||
Db 97 LGIIVAPVNDKY--IERELI--HSIGIVKPRIVFCSKNT-FQKVLN-----VK 139
|||
|||
QY 150 GELAGRRYTSFBSLLQPGGIEADNAPATGPT-----IAKPLFTSG 192
|||
|||
Db 140 SKKSIETIITIIIDLNDLGGYQCLNNFISQNSDSNLDVKKFKPYSPNRDDQVALIMFSSG 199
|||
|||
QY 193 STKLPAVPTTQRMLCANQOMLQTFPVFG---EPPVLVDMLPMNHTFGGSHNIG---- 245
|||
|||
Db 200 TTGLPRGQWMLTHKNIVA--RFSIAKOPTFGNAINPFSAILTVIPFHGFGMMTTLGFTC 257
|||
|||
QY 246 ---IVLNGSTYVLLDDGKPTAOGPAET--LRNLSEISPAVLYTPPKGMEELVGALEBDST 300
|||
|||
Db 258 GFRVVLMM-----TFEEKLFLQSLQDYKVESTLLVP-----TLMAFLKSLAL 299
|||
|||
QY 301 LRRRFRPAMKLFPPAAGLSQGIWDLDRVAEGHGERIR-----MMAIGMTEAPS 353
|||
|||
Db 300 VEKYDLSHLKEIASGAPLSKEI-----GEMVKKRPKLMFVRQGYGLTETTS 347
|||
|||
QY 354 CFTTTPPLSMAGYIGLPAPGCEVKLV-PVDGKLEG-----RFGPHVMGYWRAPBQ 405
|||
|||
Db 348 VLTTPGDAKPGSTGKIIVPFAVKVVDPTTGKILGPNRBEGLYFKGPMIMKGYNNBEAT 407
|||
|||
QY 406 AQAPEDEGYTCGSDAIKLADPADPQKGLMFDGRIAEDFKLSSGVFVSVGLRTRAVLEGG 465
|||
|||
Db 408 KAIIDNDGMLRSGLD-----AYYD-----NDGHFYVLDRLKSLIKYKG- 445
|||
|||
QY 466 SYVLDDVVAAPDECCGLVFPRLDGRALSGKESDAEVLASEVRVAMPMDMLKRLN 525
|||
|||
Db 446 ---YQVAPAEIEGI-LIQHPYIVDA-GVTGIIPDEAAGELPAAGVVVQTG-----KYL 493
|||
|||
QY 526 RE-----ATGNASRIMW-----VGLDTPPSIDKGEVTDK 555
|||
|||
Db 494 EQIVQDYVASQVSTAKMLRGVAFKFLDEIRPKGSTGKIDRK 532
|||
|||
RESULT 10
US-09-396-154-44
; Sequence 44, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; TITLE OF INVENTION: Production
; FILE REFERENCE: 341.012US1
; CURRENT APPLICATION NUMBER: US/09/396,154
; CURRENT FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: US 09/156,946
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FaastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant luciferase
US-09-396-154-44
```

```
Query Match 8.4%; Score 257.5; DB 4; Length 544;
Best Local Similarity 21.9%; Pred. No. 2,7e-18;
Matches 124; Conservative 83; Mismatches 227; Indels 131; Gaps 22;
```

```
QY 30 CVABANGEMRIRISVAEMFHNVRATNOSLLPYGLASERELLVSGNDLHLQALAFQAM 89
|||
|||
Db 38 CIALTMNHTK-ENVLYEEFLKLSGRLEASFRTKGTGLKONDTTAAVCSNSLOFPLPVIASLY 96
|||
|||
```

```
QY 90 AGIPYCPSPAYSLSLQDLAKLHIVGLQPGLVFAADAAPFORALETLLPDDVPAIFTR 149
|||
|||
Db 97 LGIIVAPVNDKY--IERELI--HSIGIVKPRIVFCSKNT-FQKVLN-----VK 139
|||
|||
QY 150 GELAGRRYTSFBSLLQPGGIEADNAPATGPT-----IAKPLFTSG 192
|||
|||
Db 140 SKKSIETIITIIIDLNDLGGYQCLNNFISQNSDSNLDVKKFKPYSPNRDDQVALIMFSSG 199
|||
|||
QY 193 STKLPAVPTTQRMLCANQOMLQTFPVFG---EPPVLVDMLPMNHTFGGSHNIG---- 245
|||
|||
Db 200 TTGLPRGQWMLTHKNIVA--RFSIAKOPTFGNAINPFSAILTVIPFHGFGMMTTLGFTC 257
|||
|||
QY 246 ---IVLNGSTYVLLDDGKPTAOGPAET--LRNLSEISPAVLYTPPKGMEELVGALEBDST 300
|||
|||
Db 258 GFRVVLMM-----TFEEKLFLQSLQDYKVESTLLVP-----TLMAFLKSLAL 299
|||
|||
QY 301 LRRRFRPAMKLFPPAAGLSQGIWDLDRVAEGHGERIR-----MMAIGMTEAPS 353
|||
|||
Db 300 VEKYDLSHLKEIASGAPLSKEI-----GEMVKKRPKLMFVRQGYGLTETTS 347
|||
|||
QY 354 CFTTTPPLSMAGYIGLPAPGCEVKLV-PVDGKLEG-----RFGPHVMGYWRAPBQ 405
|||
|||
Db 348 VLTTPGDAKPGSTGKIIVPFAVKVVDPTTGKILGPNRBEGLYFKGPMIMKGYNNBEAT 407
|||
|||
QY 406 AQAPEDEGYTCGSDAIKLADPADPQKGLMFDGRIAEDFKLSSGV--FVSVGLRTRAVL 462
|||
|||
Db 408 KAIIDNDGMLRSGLD-----LAVYDN-----DGHFYVLDRLKSLIKYKGQVAPAEIEGIL 457
|||
|||
QY 463 EGGSYVLDDVVAAPDECCGLVFPRLDGRALSGKESDAEVLASEVRVAMPMDMLKRLN 519
|||
|||
Db 458 LQHPYIVDAGVGIIPDEAAG-----ELPAAVVVVQGTGKILNEQIVQDYVASQVS 506
|||
|||
QY 520 ---MLK-----RLNRBATGNASR 534
|||
|||
Db 507 TAKMLRGVIFLDEIRPKGSTGKIDRK 531
|||
|||
```

```
RESULT 11
US-09-396-154-21
; Sequence 21, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; TITLE OF INVENTION: Production
; FILE REFERENCE: 341.012US1
; CURRENT APPLICATION NUMBER: US/09/396,154
; CURRENT FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: US 09/156,946
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FaastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant luciferase
; NAME/KEY: UNSURE
; LOCATION: (354)...(355)
; OTHER INFORMATION: Unknown amino acids
US-09-396-154-21
```

```
Query Match 8.3%; Score 255.5; DB 4; Length 544;
Best Local Similarity 22.9%; Pred. No. 4.4e-18;
Matches 132; Conservative 92; Mismatches 220; Indels 133; Gaps 27;
```


Query Match 8.2%; Score 252.5; DB 4; Length 544;
Best Local Similarity 22.3%; Pred. No. 9,2e-18;
Matches 129; Conservative 87; Mismatches 226; Indels 137; Gaps 24;

```
QY 30 CVAARANGWRRIISYAEFMHVRATIAQSLPFGLSAEPRLLVSGNDLHQLAFGAMY 89
DB 38 CIALTNHNRK-ENVLYEEFLKSCRLAESFKKYGKLNONTIAVCSENGLOFLPVIASLY 96
QY 90 AGIPYCPVSPAYSLSDIARHIVGLIQGLVPAADAAPFORAIFETIIPDVPAIFTR 149
DB 97 LGIIAIPVSDKY--IERELI--HSLGIKPRRIIFCSKNT-FQKVLN-----VK 139
QY 150 GELIAGRTVSFDSLLEPGIEADNAPATGPDT-----IAKFLFTSG 192
DB 140 SKLSKSVETIILINLEDLGGYQCINNFIQNSDSNLDVKKFKPYSFNRDQVALVMSSSG 199
QY 193 STLPRVPTPTQMLCANQOMLQTFPVFG---EEPPVLVDMPLMNTTFGSHNIG---- 245
DB 200 TTQVPPGVMLTHKNIVA--RFSIAKQDPTFGNAINPTTALITVLPFHGFGMMTTLGIFTC 257
QY 246 ---IVLYNGSTYYLDDGKPTAOGFAET--LRNLSEISPTAYLTVPKGWELVGALEERDST 300
DB 258 GFRVVLNMH-----TFEEKLFLQSLQDYKVESTLLVP-----TLMAFLAKSAL 299
QY 301 LRRERPARMKLFFPAAAGLSQGIWDRIDRVAEQHCGERIR-----MMAGLMTEETAPS 353
DB 300 VEKYDLSHLKEIASGAPLSKEI-----GEMVKKRFRKLNFRQGYGLTETTS 347
QY 354 CTFTTPELSMAGYIGLPAPCEYKLV-PVDGKLEG-----RFHGHVNSGYWRAPBEQN 405
DB 348 VLIIPKGDAPGSTGKIIPRHAKVVDPTTGKILGNPEBELYFKGDMIMKGYNNBEAT 407
QY 406 AQAFDEGYCSDAIKLDADPADPQKGLMFDGRIABDFKLSGCVFVSVGLRTAAVLEGG 465
DB 408 KAIIDNGMLRSGDI-----AYYD-----NDGHFYIVDRILKSLIKYKG- 445
QY 466 SYLDDVVAAPDEECGLVFPRLDORALSGKESDAEVLASEVRAMPFMDMLKRLN 505
DB 446 ---YQVAPAEIEGI-LIQHPYIVDA-GVTGIPDEAGELPAAGVVVQGT-----KYLN 493
QY 526 RE-----ATGNASRIW-----VGLDTPPSIDKGEVTDK 555
DB 494 EQIVQDFVSSQVSTAKMLRGVAFKFLDEIPKSGTGKIDRK 532
```

RESULT 14
US-09-396-154-20
; Sequence 20, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; FILE REFERENCE: 341.012051
; CURRENT APPLICATION NUMBER: US/09/396,154
; CURRENT FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: US 09/156,946
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant luciferase
; NAME/KEY: UNSURE

; LOCATION: (354)...(355)
; OTHER INFORMATION: Unknown amino acids
US-09-396-154-20

Query Match 8.2%; Score 251.5; DB 4; Length 544;
Best Local Similarity 22.8%; Pred. No. 1,2e-17;
Matches 131; Conservative 90; Mismatches 225; Indels 129; Gaps 26;

```
QY 30 CVAARANGWRRIISYAEFMHVRATIAQSLPFGLSAEPRLLVSGNDLHQLAFGAMY 89
DB 38 CIALTNHNRK-ENVLYEEFLKSCRLAESFKKYGKLNONTIAVCSENGLOFLPVIASLY 96
QY 90 AGIPYCPVSPAYSLSDIARHIVGLIQGLVPAADAAPFORA-----ITFIIP 140
DB 97 LGIIAIPVSDKY--IERELI--HSLGIKPRRIIFCSKNT-FQKVLNFKSKLYVETIIL 150
QY 141 DDVPAIFTEGELAGRTVSFDSLLEPGIEADNAPATGPDT-----IAKFLFTSG 196
DB 151 LDI-----NEDLGGYQCL--NNFTSQNSDINDVKKFKPYSFNRDQVALVMSSSGTGV 203
QY 197 PKAVPTPTQMLCANQOMLQTFPVFG---EEPPVLVDMPLMNTTFGSHNIG-----I 246
DB 204 PKGVMLTHKNIVA--RFSIAKQDPTFGNAINPTTALITVLPFHGFGMMTTLGIFTCGFRV 261
QY 247 VLYNGSTYYLDDGKPTAOGFAET--LRNLSEISPTAYLTVPKGWELVGALEERDSTLRR 304
DB 262 VLMH-----TFEEKLFLQSLQDYKVESTLLVP-----TLMAFLAKSALVERKY 303
QY 305 FPARMKLFFPAAAGLSQGIWDRIDRVAEQHCGERIR-----MMAGLMTEETAPSCTFT 357
DB 304 DUSHLKEIASGAPLSKEI-----GEMVKKRFRKLNFRQGYGLTETTSVLIIT 351
QY 358 TGPLSMAGYIGLPAPCEYKLV-PVDGKLEG-----RFHGHVNSGYWRAPBEQNAQAF 409
DB 352 PKXXVAPGSTGKIIPRHAKVVDPTTGKILGNPEBELYFKGDMIMKGYNNBEATKAI 411
QY 410 DEEGYCSGDAIKLDADPADPQKGLMFDGRIABDFKLSGCVFVSVGLRTAAVLEGGSYVL 469
DB 412 DKDGMRLSGDI-----AYYD-----NDGHFYIVDRILKSLIKYKG----- 445
QY 470 DVVVAAPDEECGLVFPRLDORALSGKESDAEVLASEVRAMPFMDMLKRLNREAT 529
DB 446 ---YQVAPAEIEGI-LIQHPYIVDA-GVTGIPDEAGELPAAGVVVQGT-----KYLNEQIV 497
QY 530 GN-----ASRIW-----VGLDTPPSIDKGEVTDK 555
DB 498 QNFVSSQVSTAKMLRGVAFKFLDEIPKSGTGKIDRK 532
```

RESULT 15
US-09-602-628-2
; Sequence 2, Application US/09602628
; Patent No. 6495355
; GENERAL INFORMATION:
; APPLICANT: Bames, Brian
; APPLICANT: Contag, Christopher
; TITLE OF INVENTION: Red-Shifted Luciferase
; FILE REFERENCE: SUN-127
; CURRENT APPLICATION NUMBER: US/09/602,628
; CURRENT FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: 60/140,598
; PRIOR FILING DATE: 1999-06-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Red shifted luciferase mutation
US-09-602-628-2

Query Match 8.1%; Score 249.5; DB 4; Length 550;

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2005, 21:35:24 : Search time 113 Seconds
(without alignments)
1869.837 Million cell updates/sec

Title: US-09-750-986D-30
Perfect score: 3082
Sequence: 1 MRSLEALLPERRILRLERH.....KVDAIYRGDSMLRDEATL 589

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues
Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1652.5	53.6	633	6	ADA35232 Acinetobacter
2	775	25.1	423	3	AAy87970 Pseudomon
3	716.5	23.2	336	3	AAy87971 Pseudomon
4	698.5	22.7	161	3	AAy87972 Pseudomon
5	312.5	10.1	607	5	ABP65419 Bifidobac
6	300.5	9.8	648	5	AAU44298 Propionib
7	300.5	9.8	648	6	ABM40817 Propionib
8	300	9.7	457	7	ADB74303 Mycobacte
9	298	9.7	488	7	ADB74534 Mycobacte
10	283	9.2	700	5	ADL23382 Plant acy
11	283	9.2	700	7	ADL72359 Thale cre
12	283	9.2	700	8	ADN72469 Arabidops
13	282	9.1	700	7	ADL72422 Arabidops
14	279	9.1	700	7	ADL72260 Arabidops
15	278.5	9.0	695	5	ABP66163 Bifidobac
16	278	9.0	701	5	ADL23381 Plant acy
17	278	9.0	701	7	ADL72359 Arabidops
18	275	8.9	701	7	ADL72421 Arabidops
19	266	8.6	614	7	ADP66781 Bacterial
20	265	8.6	545	2	AAy06860 Mutant lu
21	265	8.6	6842	8	AD139289 S. hygro
22	264.5	8.6	546	2	AAy06859 Mutant lu
23	263.5	8.5	544	2	AAU01228 Beetle lu
24	263	8.5	634	8	ADM86620 Human VLC
25	263	8.5	634	8	ADM16997 Human VLC

26	263	8.5	663	8	ADM16999 Human VLC
27	263	8.5	724	6	ABP57717 Novel hum
28	262.5	8.5	544	4	AAU01229 Beetle lu
29	262.5	8.5	544	4	AAU01233 Beetle lu
30	262.5	8.5	544	5	ABP53346 Photuris
31	262.5	8.5	544	5	ABP53344 Photuris
32	262.5	8.5	544	8	ADM32121 Mutant lu
33	262.5	8.5	544	8	ADM32119 Mutant lu
34	262.5	8.5	4317	6	ABU39938 Protein e
35	260	8.4	628	4	ABU3215 Human met
36	257.5	8.4	544	4	AAU01232 Beetle lu
37	257.5	8.4	544	5	ABP53345 Photuris
38	257.5	8.4	544	8	ADM32120 Mutant lu
39	256	8.3	536	6	ABU15399 Protein e
40	256	8.3	536	8	ADK13781 E. coli i
41	255.5	8.3	544	4	AAU01226 Beetle lu
42	255.5	8.3	546	2	AAy06857 Mutant lu
43	254.5	8.3	544	2	AAU01227 Beetle lu
44	254.5	8.3	546	2	AAy06858 Mutant lu
45	253	8.2	662	4	AAU39437 Propionib

ALIGNMENTS

RESULT 1
ADA35232 standard; protein; 633 AA.
XX
AC ADA35232;
XX
DT 20-NOV-2003 (first entry)
XX
DE Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
XX
DE Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KM plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PP 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
PI Breton G, Bush D;
XX
XX WPI; 2003-576092/54.
DR N-PSDB; ADA31106.
XX
PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
PS Example; SEQ ID NO 6519; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii; nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
SQ Sequence 633 AA:

Query Match	53.6%;	Score 1652.5;	DB 6;	Length 633;
Best Local Similarity	53.4%;	Pred. No. 1.9e-152;		
Matches 310;	Conservative 100;	Mismatches 168;	Indels 3;	Gaps 3;

5 EALLPFFPRILERLEHMAKTRPEQICVAARANGEMKRISVAMFNHNVRAIAOSLLPYGL 64
45 EQLKPYRQKLTDRILHFAQTKRPDHLPAKRNAGGEWVKLSYAEVLQRAWHIAQALHARNL 104

65 SAERPLLIVSGNDLEHLQLAFGAYAGIPYCPVSPAYSLSDLAKRHIVGLLQPGLYF 124
 105 SQERPLVLISGNDLEHLTISMAAMLAGVPSPAISPAYSLISQDFGKLKHVFEVLTGMY 164

125 AADAPFQRAIETLLPDVPALFTTGELAGARTVSPDLSLQPGIEADNAPATGPDI 184
 165 ASDGQAFKAIAIQACITPDIEVVTNKGIVGDICTSFQSLDTPVVS-NVQEPYQTLDENOI 223

```

185 AKFLFSSISIKLPAAVPIITQMLCAVQOMLIQTFPGEEPPVLVDWIPWNHIFGGSHI 244
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 AKFLFTSGSTKLPRAPVPTIHLMLCVNQOMLIQTFPEEEETPVLLDWISWHHTFGGSHV 283

```

243 GVLVINGSI I LDDGNFVNGFAELLNUNBLSIFIALDI VFGWGEELVGALEKUSILKRR 304
284 GLALYNGSTIYIDDGKPVAGKFDETIRNLKEISPTVYLLNVPKGMBEELTEALEKDEBELRRR 343

344 PFAKVKILFAGALSEAGNNRLDIQAQHGCEKIRIMSGMTEETAPSCAFITTPRVMA 403

404 GFIGYPAGECEIKLVPCGDKLEFCVRGKHVWKGYWRLLKADQOSTIFDEGEFFHTGDVRL 463

```

464 VVNDPTGLMYDGRIAEDFKLNTGTFFVNVGTLRNKVLIGQULLIQDYCITGSNMAIGF 523

```

```

524 LIFPKLDACQYAGUKLGEISAEILQHPKVQOMFQFLTTNNKDATGSSNNTVSMLYLMT 583

```

b
584 EPPQLDAGEVTDKGNLNQSSITKRAALIDELYQKQTDNPL 624

RESULT 2
AY87970
D AY87970 standard; protein; 423 AA.

21-SEP-2000 (first entry)

Eugenol; ferulic acid; conferyl alcohol; conferyl aldehyde; vanillin; vanillic acid; flavonring; ferulic acid-CoA-synthetase; fcs: pseudomonas sp. k199 tcs-omega-km protein.

Pseudomonas sp.

04-MAY-2000.

X
R 31-OCT-1998; 98DE-01050242.
X

X
I Rabenhorst J, Steinbuechel A, Priefert H, Overhage J;

XX
DR
XX
PT
XX
PS

WI: 2000-340642/30.
N-PSDB: AAA39720.

Transformed organisms, used for production of vanillin and other methoxyphenols, have altered catabolism of eugenol or ferulic acid.

Disclosure, Fig 2g, 80pp, German.

This invention describes novel transformed and/or mutated unit- or multi-cellular organisms (A) in which enzymes (I) involved in catabolism of eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that the organism accumulates the intermediates coniferyl alcohol (CAI), coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA). INDEPENDENT CLAIMS are also included for the following: (a) gene structures in which sequences that encode one or more of the enzymes CAI-dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-aldehyde, beta-ketothiolase, V-dehydrogenase and VA-demethylase have been altered and/or inactivated; (b) vector containing at least one gene structure of (a); (c) biotechnical production of organic compounds (particularly alcohols, aldehydes or organic acids) using (A); and (d) production method of (A). (A) are used to prepare organic compounds, preferably alcohols, aldehydes and organic acids, especially CAI, CA2, FA, V and/or VA, most particularly V, a well-known natural flavouring. The inexpensive starting material eugenol can be converted to V in a single step. This sequence represents the *Pseudomonas* sp. HR19 ferulic acid-CoA-synthetase, *fts-omega*-Km which is described in the method of the invention

Query Match	25.1%;	Score 775;	DB 3;	Length 423;
Best Local Similarity	37.4%;	Pred. No. 1.2e-66;		
Matches 231;	Conservative 43;	Mismatches 121;	Indels 222;	Gaps 25;

QY 1 MSLSLEALLPFCRILIELEHNAKTRPEQTCVAAARAANGEMRISYSYEMFHNVAIAQSLL 60

Db 1 MSLSLEALLPFCRILIELEHNAKTRPEQTCVAAARAANGEMRISYSYEMFHNVAIAQSLL 60

61 PGLSARPLILVSGNDLEHDLAFGMYAGIPIYCEPSPASLLSDLAKRIHVGLLP 120

QY	121	GVFEADAAAPFG	RAIETLPDPDPAIF	TRGELAGRRVSP	FSLLPGG	LEADNAFAATG	180
Db	121	GVFEADAAAPG	-----	MIEDG	-----	-----	138

Db 139 -----LHAGS--PAA--WVERTL-----FG-----YDMA--QQTIGC 163

```

OY      241 ISHNIGLVINGITFLD-UGKTAAGCAELLENUNSLSPALIVFNQMEELVGLLRDS    255
          |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      164 S-----DAAFRLSAQGEPVL--FVKT-----DLSGAL---N    190

```

191 ELUDE-----AARLS-W-----LATTGVPPCAVLVD 214
 Db

Db 215 VTEAGR-----DWLLTGEVPGDILLSSH-LAPAEKVSIM-----AD 250

```

251 AMRLTLDPATCFPDHQAQKRIERATRMEXGLVODDDDBEHQ---GLAPAFELPARK 307

```

```

      | : |   | : || |   | : || |   | : |   | : |
308 AKRPDGE--DLVYTHGDACLENIMVENGRSFGFIDCGRLGVADRYQDIALATRDIAEE 363

```

1. **Introduction**
 2. **Method**
 3. **Results**
 4. **Discussion**
 5. **Conclusion**
 6. **References**
 7. **Appendix**
 8. **Table 1**
 9. **Table 2**
 10. **Table 3**
 11. **Table 4**
 12. **Table 5**
 13. **Table 6**
 14. **Table 7**
 15. **Table 8**
 16. **Table 9**
 17. **Table 10**
 18. **Table 11**
 19. **Table 12**
 20. **Table 13**
 21. **Table 14**
 22. **Table 15**
 23. **Table 16**
 24. **Table 17**
 25. **Table 18**
 26. **Table 19**
 27. **Table 20**
 28. **Table 21**
 29. **Table 22**
 30. **Table 23**
 31. **Table 24**
 32. **Table 25**
 33. **Table 26**
 34. **Table 27**
 35. **Table 28**
 36. **Table 29**
 37. **Table 30**
 38. **Table 31**
 39. **Table 32**
 40. **Table 33**
 41. **Table 34**
 42. **Table 35**
 43. **Table 36**
 44. **Table 37**
 45. **Table 38**
 46. **Table 39**
 47. **Table 40**
 48. **Table 41**
 49. **Table 42**
 50. **Table 43**
 51. **Table 44**
 52. **Table 45**
 53. **Table 46**
 54. **Table 47**
 55. **Table 48**
 56. **Table 49**
 57. **Table 50**
 58. **Table 51**
 59. **Table 52**
 60. **Table 53**
 61. **Table 54**
 62. **Table 55**
 63. **Table 56**
 64. **Table 57**
 65. **Table 58**
 66. **Table 59**
 67. **Table 60**
 68. **Table 61**
 69. **Table 62**
 70. **Table 63**
 71. **Table 64**
 72. **Table 65**
 73. **Table 66**
 74. **Table 67**
 75. **Table 68**
 76. **Table 69**
 77. **Table 70**
 78. **Table 71**
 79. **Table 72**
 80. **Table 73**
 81. **Table 74**
 82. **Table 75**
 83. **Table 76**
 84. **Table 77**
 85. **Table 78**
 86. **Table 79**
 87. **Table 80**
 88. **Table 81**
 89. **Table 82**
 90. **Table 83**
 91. **Table 84**
 92. **Table 85**
 93. **Table 86**
 94. **Table 87**
 95. **Table 88**
 96. **Table 89**
 97. **Table 90**
 98. **Table 91**
 99. **Table 92**
 100. **Table 93**
 101. **Table 94**
 102. **Table 95**
 103. **Table 96**
 104. **Table 97**
 105. **Table 98**
 106. **Table 99**
 107. **Table 100**
 108. **Table 101**
 109. **Table 102**
 110. **Table 103**
 111. **Table 104**
 112. **Table 105**
 113. **Table 106**
 114. **Table 107**
 115. **Table 108**
 116. **Table 109**
 117. **Table 110**
 118. **Table 111**
 119. **Table 112**
 120. **Table 113**
 121. **Table 114**
 122. **Table 115**
 123. **Table 116**
 124. **Table 117**
 125. **Table 118**
 126. **Table 119**
 127. **Table 120**
 128. **Table 121**
 129. **Table 122**
 130. **Table 123**
 131. **Table 124**
 132. **Table 125**
 133. **Table 126**
 134. **Table 127**
 135. **Table 128**
 136. **Table 129**
 137. **Table 130**
 138. **Table 131**
 139. **Table 132**
 140. **Table 133**
 141. **Table 134**
 142. **Table 135**
 143. **Table 136**
 144. **Table 137**
 145. **Table 138**
 146. **Table 139**
 147. **Table 140**
 148. **Table 141**
 149. **Table 142**
 150. **Table 143**
 151. **Table 144**
 152. **Table 145**
 153. **Table 146**
 154. **Table 147**
 155. **Table 148**
 156. **Table 149**
 157. **Table 150**
 158. **Table 151**
 159. **Table 152**
 160. **Table 153**
 161. **Table 154**
 162. **Table 155**
 163. **Table 156**
 164. **Table 157**
 165. **Table 158**
 166. **Table 159**
 167. **Table 160**
 168. **Table 161**
 169. **Table 162**
 170. **Table 163**
 171. **Table 164**
 172. **Table 165**
 173. **Table 166**
 174. **Table 167**
 175. **Table 168**
 176. **Table 169**
 177. **Table 170**
 178. **Table 171**
 179. **Table 172**
 180. **Table 173**
 181. **Table 174**
 182. **Table 175**
 183. **Table 176**
 184. **Table 177**
 185. **Table 178**
 186. **Table 179**
 187. **Table 180**
 188. **Table 181**
 189. **Table 182**
 190. **Table 183**
 191. **Table 184**
 192. **Table 185**
 193. **Table 186**
 194. **Table 187**
 195. **Table 188**
 196. **Table 189**
 197. **Table 190**
 198. **Table 191**
 199. **Table 192**
 200. **Table 193**
 201. **Table 194**
 202. **Table 195**
 203. **Table 196**
 204. **Table 197**
 205. **Table 198**
 206. **Table 199**
 207. **Table 200**
 208. **Table 201**
 209. **Table 202**
 210. **Table 203**
 211. **Table 204**
 212. **Table 205**
 213. **Table 206**
 214. **Table 207**
 215. **Table 208**
 216. **Table 209**
 217. **Table 210**

DB 364 LGSEWADPFLVLYGIAAPDSORIAFYRLD-----EFFVLQWRSKAVD 406

QY 573 ALYRGEDQSMRLRDEATL 589

DB 407 ALYRGEDQSMRLRDEATL 423

RESULT 3

AAV87971

ID AAV87971 standard; protein; 336 AA.

XX AAV87971;

AC 21-SEP-2000 (first entry)

DT 21-SEP-2000 (first entry)

XX Pseudomonas sp. HR199 fcs-omega-Gm protein.

DE Pseudomonas sp. HR199 fcs-omega-Gm protein.

XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;

KM vanillic acid; flavouring; ferulic acid-CoA-synthetase; fcs;

KM fcs-omega-Gm.

XX Pseudomonas sp.

OS DE19850242-A1.

XX 04-MAY-2000.

XX 31-OCT-1998; 98DE-01050242.

XX 31-OCT-1998; 98DE-01050242.

XX (HAAR) HAARMANN & REIMER GMBH.

PA Rebenhorst J, Steinduechel A, Priefert H, Overhage J;

PI WPI; 2000-340642/30.

DR N-PSDB; AAA39721.

XX

PT Transformed organisms, used for production of vanillin and other

XX methoxyphenols, have altered catabolism of eugenol or ferulic acid.

XX

PS Disclosure; Fig 2h; 80pp; German.

XX

XX This invention describes novel transformed and/or mutated uni- or multi-

CC cellular organisms (A) in which enzymes (I) involved in catabolism of

CC eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that

CC the organism accumulates the intermediates coniferyl alcohol (CA1),

CC coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).

CC INDEPENDENT CLAIMS are also included for the following: (a) gene

CC structures in which sequences that encode one or more of the enzymes CA1-

CC dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-

CC aldolase, beta-ketocholase, V-dehydrogenase and VA-demethylase have been

CC altered and/or inactivated; (b) vector containing at least one gene

CC structure of (a); (c) biotechnical production of organic compounds

CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)

CC production method of (A). (A) are used to prepare organic compounds,

CC preferably alcohols, aldehydes and organic acids, especially CA1, CA2,

CC FA, V and/or VA, most particularly V, a well-known natural flavouring.

CC The inexpensive starting material eugenol can be converted to V in a

CC single step. This sequence represents the Pseudomonas sp. HR199 ferulic

CC acid-CoA-synthetase, fcs-omega-Gm which is described in the method of the

CC invention

XX

SQ Sequence 336 AA;

Query Match 23.2%; Score 716.5; DB 3; Length 336;

Best Local Similarity 35.7%; Pred. No. 4.8e-61;

Matches 211; Conservative 33; Mismatches 90; Indels 257; Gaps 19;

QY 1 MRSIALLPFGRILEHLEHMAKTRPQTCVAAAPANGEMRRISYAEHFNRAIASL 60

DB 1 MRSIALLPFGRILEHLEHMAKTRPQTCVAAAPANGEMRRISYAEHFNRAIASL 60

QY 61 PYGSAERPLLIVSGNDLEHQLAFGAMVAGIPYCPVSPAYSLLSODLAKRLHIVGLQ 120

DB 61 PYGSAERPLLIVSGNDLEHQLAFGAMVAGIPYCPVSPAYSLLSODLAKRLHIVGLQ 120

QY 121 GLVEPADAPFPORAIFETILPDVPAIFTR--GELAGRTVSFDSLLBQPGIEADNAFAA 178

DB 121 GLVEPADAPFPOMLRSS--NDVTQGSRPKTLGG-----SSMGIRF-----CR 163

QY 179 TGPPTIAKFLFTSGSTGLPRAVPTTQMLCANQOMLQTFPVGGEPPVLVDMLPMNHTF 238

DB 164 LGPDQV-----KSMRALDLFGRE-----F 183

QY 239 GGSNIGIVLYNGGTYLDDGKPTAGFAETLRNLSEISPTAYLTVPKGWEELVGAERD 298

DB 184 G--DVATY--SQHQPSDVLGMLSKTITALAAF-----DEAVVGL----- 223

QY 299 STLEREPFARMKLEFFFAAGLSQGIWDRDLVAEQHGERIRMMAGLGTETAPSCFTTT 358

DB 224 ----- 223

QY 359 GPLSMAGYIGLPARGCEVKLVVDGKLEGRFHGPHVMSGYWRAPEDONQAPEDEGYCSC 418

DB 224 --AAVY-LP-----RFEQPR-----SEIY----- 239

QY 419 DAIKLADPADPQKGLMFGRIADPFKLSGVFVSGPLRTRAVLGGSYLVDVVAAPDR 478

DB 240 -----IYDLAVSGEHR--RQGIATRLILKHEANALGAYIYVQADYGD 283

QY 479 ECIQLVFPRLDCRALSGKESASDAVLASEBRAWFADWLKRLNREATGNASRIMV 538

DB 284 PAVAL--YTKL-----GIRE-----V 298

QY 539 GLDPTPSIDKGEVTDKSGINGRAVLQWRSKAVDALYRGEDQSMRLRDEATL 589

DB 299 MHPDIDS-----TATVLQWRSKAVDALYRGEDQSMRLRDEATL 336

RESULT 4

AAV87972

ID AAV87972 standard; protein; 161 AA.

XX

AC AAV87972;

XX

DT 21-SEP-2000 (first entry)

XX

XX Pseudomonas sp. HR199 fcs-delta protein.

DE Pseudomonas sp. HR199 fcs-delta protein.

XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;

KM vanillic acid; flavouring; ferulic acid-CoA-synthetase; fcs; fcs-delta.

XX Pseudomonas sp.

OS DE19850242-A1.

XX 04-MAY-2000.

XX 31-OCT-1998; 98DE-01050242.

XX 31-OCT-1998; 98DE-01050242.

XX (HAAR) HAARMANN & REIMER GMBH.

PA Rebenhorst J, Steinduechel A, Priefert H, Overhage J;

PI WPI; 2000-340642/30.

DR N-PSDB; AAA39722.

XX

PT Transformed organisms, used for production of vanillin and other

XX methoxyphenols, have altered catabolism of eugenol or ferulic acid.

XX

PS Disclosure; Fig 2i; 80pp; German.

XX

XX This invention describes novel transformed and/or mutated uni- or multi-

CC cellular organisms (A) in which enzymes (I) involved in catabolism of
CC eugenol (B) and/or ferulic acid (FA) are inactivated in such a way that
CC the organism accumulates the intermediates confertyl alcohol (CAI),
CC confertyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
CC INDEPENDENT CLAIMS are also included for the following: (a) gene
CC structures in which sequences that encode one or more of the enzymes CAI-
CC dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-
CC aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase have been
CC altered and/or inactivated; (b) vector containing at least one gene
CC structure of (a); (c) biotechnical production of organic compounds
CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
CC production method of (A). (A) are used to prepare organic compounds,
CC preferably alcohols, aldehydes and organic acids, especially CAI, CA2,
CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
CC The inexpensive starting material eugenol can be converted to V in a
CC single step. This sequence represents the Pseudomonas sp. HR199 ferulic
CC acid-CoA-synthetase, fcs-delta which is described in the method of the
CC invention

XX Sequence 161 AA;

Query Match 22.7%; Score 698.5; DB 3; Length 161;
Best Local Similarity 92.1%; Pred. No. 9e-60; Indels 1; Gaps 1;
Matches 139; Conservative 3; Mismatches 8;

QY 1 MRSLEALPFPGRILRLHMAKTRPEQTCVAAARAANGEMRRISYAEHFNVAIAQSLL 60
DB 1 MRSLEALPFPGRILRLHMAKTRPEQTCVAAARAANGEMRRISYAEHFNVAIAQSLL 60

QY 61 PYGLSERPLLYSGNDLHQLAFGAMVAGIPYCVSPAYSLLSODLAKRIIVGLDP 120
DB 61 PYGLSERPLLYSGNDLHQLAFGAMVAGIPYCVSPAYSLLSODLAKRIIVGLDP 120

QY 121 GLVFAADAAPFORAETILPDVPAIFTRGE 151
DB 121 GLVFAADAAPFORAETILPDVPAIFTRGE 150

RESULT 5
ABP65419
ID ABP65419 standard; protein; 607 AA.

XX ABP65419;
XX
XX 19-NOV-2002 (first entry)

DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:163.

XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
XX antidiarrhetic; antibacterial; inhibitor of Salmonella; detection;
XX identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
XX rotavirus; food composition; pharmaceutical composition.

OS Bifidobacterium longum.

XX EPI227152-A1.

XX 31-JUL-2002.

XX 30-JAN-2001; 2001EP-00102050.

XX 30-JAN-2001; 2001EP-00102050.

XX (NEST) SOC PROD NESTLE SA.

XX WPI; 2002-668397/72.

XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as
XX a probe or primer for detecting and/or identifying Bifidobacterium longum
XX in a biological sample.

XX Claim 3; SEQ ID NO 163; 80pp; English.

CC The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequences given in AB081842 and AB081843, or a sequence exhibiting at
CC least 90% identity or which hybridises with the sequences given in
CC AB081842 and AB081843. Also described is a polynucleotide (II) encoding a
CC fusion protein, comprising a sequence selected from 1097 sequences given
CC in ABP65258 to ABP65354 ligated in frame to a polynucleotide encoding a
CC heterologous polypeptide. (I) has antidiarrhetic and antibacterial
CC activities, and can be used as an inhibitor of Salmonella. (I) (which is
CC a probe) is useful for the detection and/or identification of
CC Bifidobacterium longum in a biological sample. A carrier containing the
CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM I-2618) can be
CC used for preventing and/or treating diarrhoea brought about by pathogenic
CC bacteria and/or rotavirus. The carrier is a food composition selected
CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC products, ice-creams, fermented cereal based products, milk based
CC powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (I) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. AB081844 to AB081850 represent
CC Bifidobacterium related nucleotide sequences given in the Sequence
CC listing from the present invention but not mentioned further within the
CC specification. N.B. The sequence data for this patent is not represented
CC in the printed specification but is based on sequence information
CC supplied by the European Patent Office

XX Sequence 607 AA;

Query Match 10.1%; Score 312.5; DB 5; Length 607;
Best Local Similarity 23.5%; Pred. No. 5.2e-21;
Matches 142; Conservative 87; Mismatches 245; Indels 129; Gaps 21;

QY 14 IERLEHMAKTRPEQTCVAAARAANGEMRRISYAEHFNVAIAQSLLPYGLSERPLLY 73
DB 24 IERLEHMAKTRPEQTCVAAARAANGEMRRISYAEHFNVAIAQSLLPYGLSERPLLY 80

QY 74 SGNDEHQLAFGAMVAGIPYCVSPAYSLLSODLAKRIIVGLDPGLVFAAD 127
DB 81 SGNDEHQLAFGAMVAGIPYCVSPAYSLLSODLAKRIIVGLDPGLVFAAD 129

QY 128 AAPQR-----ALETLPDDVPAIFTRGEIAGR-TVETDLSLEPGGIEADNAPA 177
DB 130 RERFRLDVSXCHCPALKQILMIEGNAL--GALBGLGVTSDELEHRAVATRTD--- 183

QY 178 ATGPDTIAKFLFTSGSTKLPKAVPTTQ-----MLCANQOMLQTPPVGEEBPVLVDM 232
DB 184 -----IATIVTSSGTGNPKGAEULTHKNFVSTTISAQAL-----HEVLDDHRLRLFL 233

QY 233 PMNHFFGGSHNIGVLYNGTY-YLDDGKPTAGFAETLRNLSEISPTAYLTVPKGMBEU 291
DB 234 PLAHCFARFIQVASIASDGVGVLPDTK-----TLPLDRLSPFTYLVGPVRFEXV 286

QY 292 VGALERDSTLRERFPARMKLPFFMAAGLSOGIWDRLDRVAEQH----- 334
DB 287 YNAASHKAG-----AGWKGRLFFVKAABARVWSRKEQGEQHTFAEIAERAKETLVYR 340

QY 335 -----CGER-----IRMAAGMETETAPSCFTTGPPLSMAGY 366
DB 341 TVRGALGPRIKYVACGAPLSLDIAFPVNGIGLPMIOGIGMETETAPPAATRVTDNVIGT 400

QY 367 IGLPAPCEVKKLVVDGKLEGRFHGPHVNSGYWRABEONAQAFDEGGYCSGDAIKLADP 426
DB 401 VGOPAPGSSIR-ISDEGELQVK--GNVVRGYHNLPKTAABAFADGWLRTGDLAEI--- 454

QY 427 ADPOKGLMFDGRIAEBFKLSGVFVSVGLPRTAVLEGSSYLVUDVVAAPDEBCLGLVYF 486
DB 455 -DDEGHIVTGGIKDIIITAGGKNVSPLEBE--IAKCPVHCVVADQRPFTALV- 510

QY 487 PRLDCLRALSG-----GLGKEASDAEVLASEPVRFAFWMLKRLNBEATGNASRIWVG 540
DB 511 --TLDPESIALMLPAHGLSTETPVDRLATNAAVREIQOYVDKAN-ATYSRAESVKKFAY 567

QY 541 LDT 543
|||
Db 568 LDT 570

RESULT 6
AAU44298
ID AAU44298 standard; protein; 648 AA.
XX AAU44298;
AC

27-FEB-2002 (first entry)

Propionibacterium acnes immunogenic protein #5194.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN W0200181581-A2.

01-NOV-2001.

20-APR-2001; 2001MO-US012865.

21-APR-2000; 2000US-0199047P.

02-JUN-2000; 2000US-020841P.

07-JUL-2000; 2000US-0216747P.

(CORI-) CORIXA CORP.

Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX

WPI: 2001-616774/71.

N-PSDB; AASS9522.

Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.

PS Example 1; SEQ ID NO 5493; 106pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

Sequence 648 AA:

Query Match 9.8%; Score 300.5; DB 4; Length 648;
Best Local Similarity 23.0%; Pred. No. 8.6e-20;
Matches 151; Conservative 103; Mismatches 262; Indels 141; Gaps 28;

QY 9 PFGRR--ILRLRLHMAKT-----RPEQTCVAAA-ANGEMRISYAEHFNVAIAQS 58
|||
Db 27 PVGGENFIESHLAMRAATVANGFRP-----ATRVAGGQWIIITTAELTGRRAAGLA 81
|||
QY 59 LLPYGLSAERPL-----LIVGNDLEHLQLAFGAMAYGIPYCPVSPAYLLSODLAKR 112
|||
Db 82 FVTPLGTTEDGLRGDRISLIFAGNCEPWIADLAGMTIGV--IPV--PIYPTSTPD--QIV 136
|||
QY 113 HIYGLIQPLGVPAADAPFORALETILPDDVPAIFP-----RELAGRTVSPDSL 163
|||
Db 137 HIYTDAGVRVITTAGPRELDRIEA--RDMFGLETVILINPADQVGDHGLTVLSLEQ- 193
|||
QY 164 LEOPGIE-----ADNAFAATGPDITAKFLFTSGSTLPKAVPTTQMLCANQMLQTF 218
|||
Db 194 VRQAGVSEELQYVBERMGSCPDVVAALITYSGTTGERPGWIMSHQAALAE-----LQAL 249
|||
QY 219 PVGEERPV--LVDMLEPNNHTFGSGHNIQIVLYNGCTYYLDDKPTAQGAETLRNLSEI 276
|||
Db 250 DAEFDVTPADHSLSPFLSHALEWGMMAVIRHGLCLNTFVNPRTIS-----AMLAEV 302
|||
QY 277 SPTAVLTVPKGMELVQALERDSTLERFF--ARMKLFPPA----- 315
|||
Db 303 RPTLVFVSPRLYEQVM-----SVAREKVSDEPAKLIFEWSIRIGEMWQAOEGRRPS 356
|||
QY 316 -AAGLSQGIWDRP-----DRAVEQHCGERIRMMAGLGMET 350
|||
Db 357 VSLRAHGVADRVLVLAIRDAIGPKTVLAAGAPLKEVEFFPAACGLLVCCQYGLTEA 416
|||
QY 351 APSCTFTTGPLSNAGYIGLPAFCCEVYLVVDKLEGRFHGPHVMSGYWRAPEQNAQFD 410
|||
Db 417 SPLVSEFSPGKYKFGTAGRPLVGSQMTTTE--DGEI--LYRGPNVMKGYWRAPEXTAAAI- 472
|||
QY 411 EBGYCSGDAIKLADPADPQKGLMFDGRIADPKLSGCVSVGPLETRAVLEGSVLVD 470
|||
Db 473 EDGMLHRTDIGHI-----DEGDFLVITDRKLDIIVTGNKXISQPLE-----NSLMKD 521
|||
QY 471 -----VVVAAPDRRCGLLVFPRLDCRAL-----SGKGEASDAEVLASEPVRAMFA 518
|||
Db 522 PLFEHVALGDNRPCLTLLVYKPSLPQVEELAEKHLHTSMGPEMLRSEELAE--IRKRYA 580
|||
QY 519 DWLKRINREATGNASRIWVGLDTPSPSIDKGEVTDKGSINQRAVLQWRSAKVDALY 575
|||
Db 581 EITEKLPHQEQIRDLRVLMDEF-----TTDNGLLTPLTKVRREVEREKFEIIEVMY 632
|||
RESULT 7
ABM40817
ID ABM40817 standard; protein; 648 AA.
XX
AC ABM40817;
XX
DT 20-OCT-2003 (first entry)
XX
DE Propionibacterium acnes predicted ORF-encoded polypeptide #5493.
XX
KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine.
OS Propionibacterium acnes.
XX
PN W02003033515-A1.
XX
PD 24-APR-2003.
XX
PF 11-OCT-2002; 2002MO-US032727.
XX
PR 15-OCT-2001; 2001US-00978825.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes WJ, Benson DR, Jones R, Carter D;
PI Barth B, Valliave-Douglas J;

XX WPI; 2003-381789/36.
DR N-PSDB; ACF64451.
XX
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
XX Example 1; SEQ ID NO 5493; 1481bp; English.
PS
XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an ORF (open
CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPRO at http://wipo.int/pub/published_pct_sequences
XX
SQ Sequence 648 AA;
Query Match 9.8%; Score 300.5; DB 6; Length 648;
Best Local Similarity 23.0%; Pred. No. 8.6e-20;
Matches 151; Conservative 103; Mismatches 262; Indels 141; Gaps 28;
QY 9 PEPGR--ILERLEHMAKT-----RPEQCVARA-ANGEMRIRISYAEWFHNVRAIAOS 58
DB 27 PVGENFISHLHMFRAIVANHGFRP-----ATRRQGQWIIIRYAEIGRVAAGLARA 81
QY 59 LRPYGISARRPL-----LIVSGNDLHQLAFGANYAGIPYCPVSPANSLSLQDLAKR 112
DB 82 FVTPTGLTDEGLQGRDRISLFAGNCEPWEIADLAGMTIGV--IPV-PIYPTSTPD--QIV 136
QY 113 HTVGLLQPGVAFADAAPFORALETILPPDVPAIFP-----RGLLAGRTVSPDSL 163
DB 137 HAVTDAGVAVITTAGKELEDRIEA--RDMPELETVILINPADQYGDHDLTVLSLEO- 193
QY 164 LEQPGGIE-----ADNAFATGPDITAKFLTSGSTKLPRVAVPTTORMCANQOMLLQTF 218
DB 194 VQAGVSEERIQTVBERMQSCPDVAALITSGTGEPGCVWISIRALAE-----LQAL 249
QY 219 PVFGEPRPV--LVDMLPMMHTFGSGSHNIGIVLNGSTYYIDGKPPAQAQFAETLRNLSEI 276
DB 250 DAFFDVTAPDHSLSFRLPSHLEWMSMAVIRHGCINTFVNPNTKIS-----AMLALEY 302
QY 277 SPTAVLYTVKGMELVGALEDRSTLEBRF--ABMKLPFEFA----- 315
DB 303 RPTLFVSVKLYHQV-----SVAREKVS DSPAKUKITFEWSRRIRGEMWQAEQGRPRS 356
QY 316 -AAGLSQGIWDRY-----DRVAEQHCGERIRMAAGLQMTET 350
DB 357 VSLRAHGVADRVLKAIKDAIGRPKTVLAAGGAPLRKEVEBFPAACGLVCGGYGLTEA 416
QY 351 AASCTTTGTPRLSMAGYIGLPAPGCEVAVLVVDGKLEGRFRHGHVMSGYWRAPRQNAQAPD 410

DB 417 SPLVSNFSPGKYKFGTAGRPVVGSCQMTTTE-DGEI--LYRGPVVMGYMKAPATAAAI- 472
QY 411 EEGYCSGPAIKLADPADPOKGMFDPRIAEPEKLSGVFVSGPRLTRVAVLEGSGYVD 470
DB 473 EDQWLHTGIGHI-----DEDFGLVITDRKLDIIVTLANGNISQPIE-----NSLMD 521
QY 471 -----VVVAAPPRECVGLLVFPRLLDCRAL-----SGLGKASDAEVLASBPVAMPRA 518
DB 522 PLFEHAVLLGDNRPCLTLLVVKPSLPQVEELAEERLHTSMTGPEMLRSEELAE-IRRYVA 580
QY 519 DMLKLRNRAATGNASIMVGLDTPSPIDKGEVYDPKSGINGRAVLQMSAKYDALY 575
DB 581 ETTEKLPHEQIQRDLRVLDEF-----TTDNGLLPTTLVRRREVRKFTETVEEMY 632
RESULT 8
ADB74303
ID ADB74303 standard; protein; 457 AA.
XX
XX ADB74303;
AC
XX 04-DEC-2003 (first entry)
DT
XX
XX Mycobacterium leprae non-naturally occurring peptide #26.
DE
XX
XX Non-naturally occurring peptide; anion pump protein; tuberculosis;
KM hypersensitivity reaction; tuberculostatic.
XX
XX Mycobacterium leprae.
OS
XX
XX US6583266-B1.
PN
XX 24-JUN-2003.
PD
XX
XX 16-SEP-1994; 94US-00311731.
PF
XX
XX 19-AUG-1993; 93US-00109181.
PR
XX 22-OCT-1993; 93US-00142558.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX Smith DR, Mao J;
PI
XX
XX WPI; 2003-656441/62.
DR N-PSDB; ADB74275.
DB
XX New Mycobacterium tuberculosis anion pump peptide useful for as
PT tuberculosis vaccine and diagnosis of tuberculosis infection.
XX
XX Disclosure; SEQ ID NO 52; 26bp; English.
PS
XX The invention relates to a non-naturally occurring peptide of
CC Mycobacterium tuberculosis comprising an amino acid sequence
CC corresponding to an anion pump protein. The invention also relates to a
CC non-naturally occurring nucleic acid corresponding to a DNA sequence of
CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
CC leprae or for screening for new tuberculosis drugs. Purified proteins
CC derived from the sequences of the invention may elicit a specific immune
CC response. The peptide may also be used to detect hypersensitivity
CC reactions of individuals exposed to Mycobacterium tuberculosis or
CC Mycobacterium leprae. The proteins and peptides may be affixed to solid
CC supports to detect antibodies typical of hypersensitivity reactions, from
CC a patient's sera. This sequence represents Mycobacterium leprae non-
CC naturally occurring peptide of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 457 AA;
Query Match 9.7%; Score 300; DB 7; Length 457;

Best Local Similarity 26.2%; Pred. No. 5.7e-20;
Matches 143; Conservative 67; Mismatches 210; Indels 126; Gaps 27;
QY 44 SYAEFMFNVAIAQSLPFGISAEKPLIYSGNDLEHLQAFGAMVYGPYCPVSPAYSL 103
DB 13 SVAEKVGARLVAVLATP---TASTVLATP-----GCLTIGVVPVPADIGV 57
QY 104 LSQDLAKLRHIV-----GLQPGLVFAADAPFQRAIETILPDDVAIFRGLAGRT 157
DB 58 VER-----RMLLDSGAQAWLGGP--SADSA-----PDGLPHI-----P 90
QY 158 VSPDSLLEQPGIEADNAFAATGPDITAKFLFSGSTKLPRKAVPTTQRMICANQMLLOT 217
DB 91 VQDADA-----RSNMRYPEPSPDATAMVITYSGTGPCKGVLSRAIYVDLALAQA 142
QY 218 FVPFGEEPVLVDMLPNNHTFG-----GSHNIGIYLVNGSTYYLDGCKPTAGFAETLR 271
DB 143 WQMTAAD--VLVHGLPLFHVHGLVLGLLSLRI-----GNRFVHTGKPTPTAYQA-- 191
QY 272 NLSEISPTAYLTVPKGMEEVLGALERDSTLEREFARMKLEFFAAAGLSQIMDLRVA 331
DB 192 -CEBAGSLYFGVPTWMSRLVA---DEAV-ARALRPARLLVSGSASLPVPVFDRLAHLT 245
QY 332 EOHGGERIRMMAGLGMTETAPSCFTTGPLSMAGYIGLPAPGCEVKL-----VPYDGK 384
DB 246 GHRPIER-----YGSTESLITLSTLADGERRAGWGLPLAGVQTRLVDSEGGVPYDGE 299
QY 385 LEBRF--HGFHWMSGYWRAPENQAQAFDESGYCSGPAITLADPADQKGLMFGDRAED 442
DB 300 TVGRLOYRSPTMGGYLNREPEATAEAFDEGMYRTGD--VAVDSGGMHR--IYGRSVD 355
QY 443 FKLSGVFVSVGLRTAVLEGSSVYLDV--VVAAPRECLGLVFPRLDORALSGLKE 501
DB 356 LKLGGRIGAGET--EMALLGHPDVRVAVGLPD--EDLQGRIVAFV-----VGAE 404
QY 502 ASDAEVLASBPVAMFADWMLKRLNREATGNASRIWVGLDTPPSIDKEVTDKSINOR 561
DB 405 ALDADELIN-----YVAQQLSIHKRPRE-----VRFDALPA--TRMERSKSSCSSR 450
QY 562 AVTLQMR 567
DB 451 A--DMR 454
RESULT 9
ID ADB74534 standard; protein; 488 AA.
AC ADB74534;
XX 04-DEC-2003 (first entry)
DT
XX Mycobacterium leprae non-naturally occurring peptide #234.
DE
XX Mycobacterium leprae non-naturally occurring peptide #234.
KM Non-naturally occurring peptide; anion pump protein; tuberculosis;
XX hypersensitivity reaction; tuberculostatic.
OS Mycobacterium leprae.
XX
XX US6583266-B1.
PN
XX 24-JUN-2003.
PD
XX
XX 16-SEP-1994; 94US-00311731.
PF
XX 19-AUG-1993; 93US-00109181.
PR 22-OCT-1993; 93US-00142558.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX Smith DR, Mao J;
PI
XX WPI; 2003-656441/62.
DR

XX New Mycobacterium tuberculosis anion pump peptide useful for as
PT tuberculosis vaccine and diagnosis of tuberculosis infection.
PS Disclosure; SEQ ID NO 283; 26pp; English.
XX
XX The invention relates to a non-naturally occurring peptide of
CC Mycobacterium tuberculosis comprising an amino acid sequence
CC corresponding to an anion pump protein. The invention also relates to a
CC non-naturally occurring nucleic acid corresponding to a DNA sequence of
CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
CC leprae or for screening for new tuberculosis drugs. Purified proteins
CC derived from the sequences of the invention may elicit a specific immune
CC response. The peptide may also be used to detect hypersensitivity
CC reactions of individuals exposed to Mycobacterium tuberculosis or
CC Mycobacterium leprae. The proteins and peptides may be affixed to solid
CC supports to detect antibodies typical of hypersensitivity reactions, from
CC a patient's sera. This sequence represents Mycobacterium leprae non-
CC naturally occurring peptide of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 488 AA;
Query Match 9.7%; Score 298; DB 7; Length 488;
Best Local Similarity 26.4%; Pred. No. 9.9e-20;
Matches 135; Conservative 61; Mismatches 196; Indels 120; Gaps 24;
QY 44 SYAEFMFNVAIAQSLPFGISAEKPLIYSGNDLEHLQAFGAMVYGPYCPVSPAYSL 103
DB 50 SVAEKVGARLVAVLATP---TASTVLATP-----GCLTIGVVPVPADIGV 94
QY 104 LSQDLAKLRHIV-----GLQPGLVFAADAPFQRAIETILPDDVAIFRGLAGRT 157
DB 95 VER-----RMLLDSGAQAWLGGP--SADSA-----PDGLPHI-----P 127
QY 158 VSPDSLLEQPGIEADNAFAATGPDITAKFLFSGSTKLPRKAVPTTQRMICANQMLLOT 217
DB 128 VQDADA-----RSNMRYPEPSPDATAMVITYSGTGPCKGVLSRAIYVDLALAQA 179
QY 218 FVPFGEEPVLVDMLPNNHTFG-----GSHNIGIYLVNGSTYYLDGCKPTAGFAETLR 271
DB 180 WQMTAAD--VLVHGLPLFHVHGLVLGLLSLRI-----GNRFVHTGKPTPTAYQA-- 228
QY 272 NLSEISPTAYLTVPKGMEEVLGALERDSTLEREFARMKLEFFAAAGLSQIMDLRVA 331
DB 229 -CEBAGSLYFGVPTWMSRLVA---DEAV-ARALRPARLLVSGSASLPVPVFDRLAHLT 282
QY 332 EOHGGERIRMMAGLGMTETAPSCFTTGPLSMAGYIGLPAPGCEVKL-----VPYDGK 384
DB 283 GHRPIER-----YGSTESLITLSTLADGERRAGWGLPLAGVQTRLVDSEGGVPYDGE 336
QY 385 LEBRF--HGFHWMSGYWRAPENQAQAFDESGYCSGPAITLADPADQKGLMFGDRAED 442
DB 337 TVGRLOYRSPTMGGYLNREPEATAEAFDEGMYRTGD--VAVDSGGMHR--IYGRSVD 392
QY 443 FKLSGVFVSVGLRTAVLEGSSVYLDV--VVAAPRECLGLVFPRLDORALSGLKE 501
DB 393 LKLGGRIGAGET--EMALLGHPDVRVAVGLPD--EDLQGRIVAFV-----VGAE 441
QY 502 ASDAEVLAS-----EPVAMPADMLK 523
DB 442 ALDADELINVAQQLSIHKRPREVRFDALPR 473
RESULT 10
ID ADL23382 standard; protein; 700 AA.
AC ADL23382;
XX
XX

DT	20-MAY-2004	(first entry)
XX		
DE	Plant acyl-coenzyme A synthetase amino acid sequence #6.	
XX		
KV	plant; acyl-coenzyme A synthetase; acyl-CoA synthetase; enzyme;	
KW	transgenic plant.	
XX		
OS	Unidentified.	
XX		
PN	WO200209295-A2.	
PD		
XX		
PE	31-JAN-2002.	
PP	19-JUL-2001; 2001WO-US022774.	
XX		
PR	21-JUL-2000; 2000US-0220474P.	
PR	16-JUL-2001; 2001US-00906419.	
XX		
PA	(SHOC/) SHOCKEY J M.	
XX	(SCHN/) SCHNURR J.	
XX	(BROW/) BROWSE J A.	
XX		
P1	Shockey JM, Schnurr J, Browse JA;	
DR	WPI; 2002-241594/29.	
DR	N-P8DB; ADL23371.	
PT	Novel acyl coenzyme A synthetases gene useful for altering a phenotype of a plant, making a transgenic plant and for producing variants of acyl-CoA synthetases.	
PT		
PS	Claim 2; SEQ ID NO 17; 15pp; English.	
XX		
CC	The invention comprises the amino acid and coding sequences of plant acyl coenzyme A synthetase (acyl-CoA synthetase) enzymes. The DNA and protein sequences of the invention are useful for altering a phenotype of a plant (transgenic plant). The DNA and protein sequences of the invention are also useful for producing variants of acyl-CoA synthetases. The present invention.	
CC		
CC		
CC		
SQ	Sequence 700 AA;	
	Query Match 9.2%; Score 283; DB 5; Length 700;	
	Best Local Similarity 22.5%; Pred. No. 5,1e-18;	
	Matches 150; Conservative 97; Mismatches 214; Indels 206; Gaps 35;	
OY	9 PEPGGILRLLEHMAKTRPEQCVAARAA-----GEMRKISYAEMHNRAINQSILPYGL 64	
Dd	: : : : : : : : : : : : : : : : : : : : : : : : :	
	81 PEIGTLDHNFVAHVATVAENKYLTGRVSDDGTIGBYSMWTGGEASERQAIGSGILFFHG 140	
OY	65 S-----AERPLLIIVSNGNDLEHLQAFGAMYGIFPCPVSPAYSLSLOPLAK--LRH 113	
Dd	: : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
	141 NQGDCVGLYFINRPWMVLV-----VDIACAAYSVSV-----PLYDLGPNAVVFVN 188	
OY	114 -----IVGLQLQGLVFAADAPFOARIETI--LPDVPALIFRGELAGRTVSPDSL 163	
Dd	: : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
	189 ANLQAIFCVPOQTNLITSLFLAE-IPSIIRLVVVGGADEHLPSL--TGVTIVSYOKL 244	
OY	164 LEQPGGIDADNAFAATGPDITAKLFTEGSSTGLPRAVFTTQMCLCAN---QQMLLOTTPV 220	
Dd	: : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
	245 LSQ--GRSSLHPFSPPKEDIAITICTYTGTGTTPKGVALTHGNLLANAVAGSSVEAEFP 301	
OY	221 FGEBPVAVDMVMNMH-----TRGSGHNIGVIYNGSTYYLDDCKPFAQGFAYEL 270	
Dd	: : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
	302 ----SDVYISYLPFLAHYIERANQINGVYG--VAVGFYQGVFKLM----- 342	
OY	271 RNLSEISTPAYLVTVPKGWEEVLGALERDSTLEREFPAAMKLFFFAAGLSQGI----- 323	
Dd	: : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
	343 -DFAVLARPTIFCSVRILNRIYDI--TSAVASSGVKKRLFELIYNKKQAIMNGRTPS 399	
OY	324 --WDRL--DRVAAQHCGGERIRMA-----GLGHTETAPECT 355	
	: : : : : : : : : : : : : : : : : : : : : : : : : : : : :	

Db	400	AFWMDLVENFKIKK-LGGRVRFMSGGASPLSPDVMDFLRICFGCSVRBGVGMTET--SCV	456
Qy	336	FTT---GFLSMAGVIGLPAAPGCEVVLVYVD-----GILEGFGHGHVMSGYW	399
Db	457	ISANDDGD-NLSGHGSPNPACFVVLVVPENMYTSDQPYRGEICVR--GPILFGXY	513
Qy	400	RAPEONMAAPFEEGGVYCGSDAIKLADPADDPQKGLMFDG-----RIADPFLSSGVFVS	452
Db	514	KDEQTRTEILDQDGLMHTGDI-----GLWLPGRRLKIIPRKNIIFPLAGEEYIA	562
Qy	448	RLLD--CRALSGLGKEA-----SDAEVLASEP-----VRAMPADW	520
Db	623	RVRKTVLAEMDQDLGREAQLRGFEPAKAVTLVPEPFTLENGLLTPTFKIKRPOAKAYFAEA	682
Qy	521	LKRLURE	527
Db	683	ISKMYAE	689
Db	683	ISKMYAE	689
RESULT 11			
ID	ADL72359		
ID	ADL72359	standard; protein; 700 AA.	
AC	ADL72359;		
DT	20-MAY-2004	(first entry)	
XX			
XX		Modified Arabidopsis thaliana acyl-CoA synthetase 3B protein.	
XX		plant; acyl-CoA synthetase; soybean; sunflower; cotton; maize; castor;	
KM		transgenic plant; triacylglycerol biosynthesis; fatty acid; seedling;	
KW		beta-oxidation cycle substrate; jasmonic acid; plant defence.	
XX			
OS		Arabidopsis thaliana.	
XX			
PN	WO2003087321-A2.		
XX			
PD	23-OCT-2003.		
XX			
PF	09-APR-2003; 2003WO-US010754.		
XX			
PR	09-APR-2002; 2002US-00119136.		
XX			
PR	08-APR-2003; 2003US-00119136.		
XX			
PA	(UTWA-) UNIV WASHINGTON STATE RES FOUND.		
XX			
PI	Shockey JM, Schnurr J, Browne JA;		
XX			
DR	WPI:2003-853948/79.		
XX			
DR	N-PSDB; ADL72353.		
XX			
PT		New plant acyl-CoA synthetase protein derived from soybean, sunflower,	
PT		cotton, maize, and castor, useful in cuticle was synthesis, and in the	
PT		synthesis of jasmonic acid which is involved in reproduction and plant	
PT		defense.	
XX			
XX			
PS	Claim 29; SEQ ID NO 130; 226bp; English.		
XX			
CC		The invention relates to a new purified plant acyl-CoA synthetase protein	
CC		comprising at least one of the motifs selected from 9 fully defined	
CC		motifs given in the specification, and derived from a crop plant selected	
CC		from soybean, sunflower, cotton, maize, and castor. The purified plant	
CC		acyl-CoA synthetase protein comprises at least one motif selected from: V	
CC		-P/T-L/I-Y-D/A/S-T/S-L-G; I-M-C-Y/F/K-T-S-G-T/S-T/S-G-X1-P-K-G-V; S/A-	
CC		-Y/M/F-P-L/S-A/M-H; L/O-K/R-P-T/P/S; S/G/V/-G-A/G/S-N/L/S-P-L/I/M; G-Y-G-	
CC		-L/M-T-E-T/S; P/S/A-R/K-G/G-A-B/I-V-C/K/V-I/V/L-R/G-G; IIDRKK; and L-	
CC		L/V/M/L-P/A-T/A/S-F/L/M/Y-K-X1-K/R-R. The nucleic acid is useful in	
CC		producing a transgenic plant (claimed). The plant Acyl-CoA synthetase is	
CC		useful in TAG biosynthesis, for activating fatty acids released from oil	

CC bodies in newly germinated seedlings, as substrates for the beta-oxidation cycle which supplies the plant with cellular energy until it becomes photosynthetically competent, in cuticle was synthesizing and in the synthesis of jasmonic acid, a fatty acid-derived signaling compound involved in reproduction, plant defence, and other response reactions. CC This sequence corresponds to a modified ACS3B protein of the invention.

XX Sequence 700 AA:

Query Match 9.2%; Score 283; DB 7; Length 700;
Best Local Similarity 22.5%; Pred. No. 5.1e-18;
Matches 150; Conservative 97; Mismatches 214; Indels 206; Gaps 35;

```

QY 9 PPGRIERLEHNAKTRPEOTCVAARAAN-----GEMRISYAEFMHVRALIAQSLPYGL 64
DB 81 PEIGTLHDNFVAHYEAENKYLIGTRVRSDDTIGESWMTYGEAASERQAIISGLLPHGV 140
QY 65 S-----AERPLIVSGNDLEHLQAFGAMYAGIPYCPVSPAYSLLSQDLAK--LRH 113
DB 141 NQGDGCVGLYFINRPEWLV-----VDHCAAYSFVS-----PLYDTLPDAVKKPVVNH 188
QY 114 -----IVGLIQGLVPADAAFPORALETI--LPDDVPAIFTRGELAGRTVSFDSL 163
DB 189 ANLOAIFCVQOTINILISFLAE--IPISIRLIVVVGADENHPSL--PG--TGVTVISYOKL 244
QY 164 LEOPGIEADNAPATGPDITIAKFLFTSGSTKLPKAVPTTORMICAN---QOMLQTFPV 220
DB 245 LSG--GRSSLHPPSPKPEBIDATICTSGTTPKGVLTGMLIANVAGSSVEAEFFP- 301
QY 221 FGSEPPVLVDMLEPMNH-----TFGSHNIGIVLYNGGTYVLDGKPTAQGFATL 270
DB 302 ----SDVIYSYPLAHYERANOIMGVYGG---VAAGFYQGDVFKLMD----- 342
QY 271 RNLSISPTAVLYTPKGMEEVLGALERDSTLRERFFARMKLPFAAAGLSQGI----- 323
DB 343 -DEAVLRPTIFCSPVRLYNYIDGI--TSAVKSGVYKKEFLIAYNSKKQAIINGRTPS 399
QY 324 --WDL--DRVAROHGGERIRMA-----GIMTETAPASCT 355
DB 400 AFMDKLVFNKIKER--LQGRVRFMGSGASPLSPVDMFLRICFGCSVREGYGMET--SCV 456
QY 356 FTI---GPLSMAGYIGLPARCEVKLVVD-----GKLEGRFHGPHVNSGYW 399
DB 457 ISMDGDD--NLSGVSGSPNACEVKLVDEPMNYTSDQYPRGELCVR--GPIIFGY 513
QY 400 RAEQNAQADEFDEGYCCSGDAIKLADPADPOKIMFDG-----RIADPKLSSGVFVS 452
DB 514 KBEQOTREIIDGDMHTGDI-----GIMLEFGRLKIIRKKNIFKLAQGEYIA 562
QY 453 VGPLR-----TRAVLEGSY--VLDVVVAAP-----RECGLLV-FP 487
DB 563 PEKIEENVYTKCRFVSOCTIHGDSFNSSLVAIVSVPEVMKMAASGKIKYEHIGQLNDP 622
QY 488 RLID--CRALSGLKEA-----SDAEVLASER--VRAMFADW 520
DB 623 RVAKTVLAEMDDIGREHQALGFEPKAVTLVPERFTLENGLLPTFKIKRPQAKAYAAEA 682
QY 521 LKRLNR 527
DB 683 ISKRYAB 689

```

RESULT 12
ADN72469
ID ADN72469 standard; protein; 700 AA.
XX ADN72469;
XX
DT 15-JUL-2004 (first entry)
XX
DE Thale cress protein upregulated in E2fa/Dpa expressing plants SeqID 364.
XX
KM plant; transgenic; E2fa/Dpa transcription factor; growth regulator;

KM animal feed product; thale cress; cell wall biosynthesis;
KM nitrogen metabolism; carbon metabolism.

XX Arabidopsis thaliana.

XX WO2004035798-A2.

XX 29-APR-2004.

XX 20-OCT-2003; 2003WO-BE011658.

XX 18-OCT-2002; 2002EP-00079408.

XX (CROP-) CROPDISEIGN NV.

XX Inze D, De Veylder L, Vlieghe K;

XX WPI; 2004-348466/32.

XX N-PSDB; ADN72468.

PT Altering plant characteristics, useful for producing plants for enzyme or pharmaceutical production comprises modifying in a plant, expression of one or more nucleic acids and/or modifying level or activity of one or more proteins.

PS Claim 1; SEQ ID NO 364; 134pp; English.

CC This invention relates to a novel method for altering one or more plant characteristics. Specifically, it refers to identifying genes that are up- or down-regulated in transgenic plants overexpressing the heterodimeric E2fa/Dpa transcription factor of Arabidopsis and using these sequences to alter plant characteristics accordingly. The present invention describes generating transgenic plants for the production of growth regulators, enzymes, therapeutics, pharmaceuticals and animal feed products, where the altered plant characteristics are selected from increased yield or biomass, enhanced survival capacity, stress tolerance, plant architecture or physiology, altered endoreduplication, biochemistry, signal transduction, storage lipid mobilization and/or altered photosynthesis, each relative to the corresponding wild type plants. Accordingly, these sequences can also be useful as positive or negative selectable markers during transformation of cells or tissues. The identified genes play a role in a variety of biological processes such as DNA replication, cell wall biosynthesis, nitrogen and/or carbon metabolism or they function as transcription factors. This polypeptide sequence is thale cress protein expressed by a gene upregulated 1.3 fold or more in plants overexpressing the E2fa/Dpa transcription factor, given in an exemplification of the invention.

XX Sequence 700 AA:

Query Match 9.2%; Score 283; DB 8; Length 700;
Best Local Similarity 22.5%; Pred. No. 5.1e-18;
Matches 150; Conservative 97; Mismatches 214; Indels 206; Gaps 35;

```

QY 9 PPGRIERLEHNAKTRPEOTCVAARAAN-----GEMRISYAEFMHVRALIAQSLPYGL 64
DB 81 PEIGTLHDNFVAHYEAENKYLIGTRVRSDDTIGESWMTYGEAASERQAIISGLLPHGV 140
QY 65 S-----AERPLIVSGNDLEHLQAFGAMYAGIPYCPVSPAYSLLSQDLAK--LRH 113
DB 141 NQGDGCVGLYFINRPEWLV-----VDHCAAYSFVS-----PLYDTLPDAVKKPVVNH 188
QY 114 -----IVGLIQGLVPADAAFPORALETI--LPDDVPAIFTRGELAGRTVSFDSL 163
DB 189 ANLOAIFCVQOTINILISFLAE--IPISIRLIVVVGADENHPSL--PG--TGVTVISYOKL 244
QY 164 LEOPGIEADNAPATGPDITIAKFLFTSGSTKLPKAVPTTORMICAN---QOMLQTFPV 220
DB 245 LSG--GRSSLHPPSPKPEBIDATICTSGTTPKGVLTGMLIANVAGSSVEAEFFP- 301
QY 221 FGSEPPVLVDMLEPMNH-----TFGSHNIGIVLYNGGTYVLDGKPTAQGFATL 270
DB 302 ----SDVIYSYPLAHYERANOIMGVYGG---VAAGFYQGDVFKLMD----- 342

```


AC ADL72260;
 XX
 XX 20-MAY-2004 (first entry)
 DE Arabidopsis thaliana acyl-CoA synthetase 3B protein.
 XX
 KM plant; acyl-CoA synthetase; soybean; sunflower; cotton; maize; castor;
 KM transgenic plant; triacylglycerol biosynthesis; fatty acid; seedling;
 KM beta-oxidation cycle substrate; jasmonic acid; plant defence.
 XX
 OS Arabidopsis thaliana.
 PN W02003087321-A2.
 XX
 XX 23-OCT-2003.
 PD 09-APR-2003; 2003WO-US010754.
 XX
 PF 09-APR-2002; 2002US-00119136.
 PR 08-APR-2003; 2003US-00119136.
 XX
 PA (UYMA-) UNIV WASHINGTON STATE RES FOUND.
 PI Shockley JM, Schnurr J, Browne JA;
 XX
 XX WPI: 2003-853948/79.
 DR N-PSDB: ADL72249.
 XX
 XX
 PT New plant acyl-CoA synthetase protein derived from soybean, sunflower,
 PT cotton, maize, and castor, useful in cuticle was synthesis, and in the
 PT synthesis of jasmonic acid which is involved in reproduction and plant
 PT defense.
 PS Disclosure; SEQ ID NO 17; 226pp; English.
 XX
 XX The invention relates to a new purified plant acyl-CoA synthetase protein
 CC comprising at least one of the motifs selected from 9 fully defined
 CC motifs given in the specification, and derived from a crop plant selected
 CC from soybean, sunflower, cotton, maize, and castor. The purified plant
 CC acyl-CoA synthetase protein comprises at least one motif selected from: V
 CC -P/T-L-I-Y-D/A/S-T/S-L-G; I-M-C-Y/F/K-T-S-G-T/S-T/S-G-XI-P-K-G-V; S/A-
 CC -Y/M-F-L-P-L/S-A/W-H; L/Q-K/R-P-T/P/S; S/G/V/-G-A/G/S-A/L/S-P-L/I/W; G-Y-G
 CC -L/M-T-E-T/S; P/S-A/R-K-G/A-E/I-V-C/K/V-I/V/L-R-G-G; IIPKK; and L-
 CC -L/V/M/L-A-T/A/S-F/L/M/Y-K-XI-K/R-R. The nucleic acid is useful in
 CC producing a transgenic plant (claimed). The plant Acyl-CoA synthetase is
 CC useful in TAG biosynthesis, for activating fatty acids released from oil
 CC bodies in newly germinated seedlings, as substrates for the beta-
 CC oxidation cycle which supplies the plant with cellular energy until it
 CC becomes photosynthetically competent, in cuticle was synthesis, and in
 CC the synthesis of jasmonic acid, a fatty acid-derived signaling compound
 CC involved in reproduction, plant defence, and other response reactions.
 CC This sequence corresponds to an ACS protein of the invention.
 CC
 XX Sequence 700 AA;
 SQ
 Query Match 9.1%; Score 279; DB 7; Length 700;
 Best Local Similarity 22.7%; Pred. No. 1.3e-17;
 Matches 150; Conservative 100; Mismatches 216; Indels 196; Gaps 35;
 QY 9 PFGRIILERLEHMAKTRPEQTCVAAARAAN---GEMRISYAEMFHNVRATAGSLPYGL 64
 DB 81 PEGTGLHNDVNAVETVAENKYLCTRVRSDDTIEYGMNTYGEAASERQAGSLPLRHGV 140
 QY 65 S-----AERLLIVSGNDEHLQALGAMAYAGIPYCVSPAYSLISODLAK--LRH 113
 DB 141 NOGACVGLYFINREMLV---VDHACAAVSFVSV-----PLYDTLGDADAVKFFVNH 188
 QY 114 -----IYGLDQGLVFAADAPFORALETI--LPDDVPAITRGLGRVVSFDSL 163
 DB 189 ANLOATYICVQPTLNLISFLAE-IPSTRLLVVGADENHLSL-PRG--TGVTVVSYQKL 244
 QY 164 LEOPGGIEADNAPATGPDITAKFLFTSGSTKLPRKAVPTTQRMLCAN---QOMLLQTFPV 220

DB 245 LSG--GRSSLHPFSPKPEDIATICTSGTGTGPKGVLTGNLIANVAGSSVEAEFFP- 301
 QY 221 FGEPPPLVNDMLPMNHFFGSGHNIGVLYNGCTYYLDDGKPTAGF-----AETLRNISE 275
 DB 302 ---SDVIYISYLPAAHYERANOI-MGVYGG-----VAAGFYQGDVLKLMDDPAV 346
 QY 276 ISPTAVLTVPKGMDELVALERDSTLRERFFARMKLFFFAAAGLSOGI-----WDR 326
 DB 347 LRPTIFGCVPRVLYNRITDGT--TSAVSSGVKKRLIEIAYNSKKOAIINRTSAPFMDK 404
 QY 327 L--DRVAEQHCEIRIRMA-----GLGNTETAPSGCTFTT-- 358
 DB 405 LVFNKIKEX-LGGRVRFMGSGASPLSPDVMDFLRICFGSVREGYGTET--SCVISAMD 461
 QY 359 -GPLSMAGYIGLPAFGCEVULVPD-----GLKEGFHGRPHVSGVWRAPQ 404
 DB 462 DGD-NLSGHVSGPVPACVXLVDPEMNYSSEDPYRGEICVR--GPITFKGYKDEQ 518
 QY 405 NQAQFDEGGYCSGDALKLADPADPQKGLMPDG-----RIADPFLKSGVFSVGBLR 457
 DB 519 TREILDGDMHTGDI-----GLWLPGRRLKIIDRKXIFFLAQGEYIAPEKIE 567
 QY 458 -----TRAVLEGGSY--VLDVVVAAPD-----RECLGLLV-PRRLD- 491
 DB 568 NVYTKCRFVSQCFIHGDSFNSLVAIVSDPEVKKMAASBGIKYEHILGQLCNDPRVRKT 627
 QY 492 -CRALSLGKEA-----SDAEVLASEP-----VRAMPALMKRLN 525
 DB 628 VLAEMDLGRBAQURGEFAKAVTLVPEPTLENGLLTPPFKIRPOAKAYFAEAIKMY 687
 QY 526 RE 527
 DB 688 AE 689
 RESULT 15
 ABP66163
 ID ABP66163 standard; protein; 695 AA.
 AC
 AC ABP66163;
 DT 19-NOV-2002 (first entry)
 XX
 DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:907.
 XX
 XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
 KM antidiarrhetic; antibacterial; inhibitor of Salmonella; detection;
 KM identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
 KM rotavirus; food composition; pharmaceutical composition.
 XX
 OS Bifidobacterium longum.
 XX
 XX EPI227152-A1.
 PN
 PD 31-JUL-2002.
 PD
 PF 30-JAN-2001; 2001EP-00102050.
 PR
 PR 30-JAN-2001; 2001EP-00102050.
 PA (NEST) SOC PROD NESTLE SA.
 XX
 XX WPI: 2002-668397/72.
 DR
 XX
 XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as
 PT a probe or primer for detecting and/or identifying Bifidobacterium longum
 PT in a biological sample.
 PT
 PS Claim 3; SEQ ID NO 907; 80pp; English.
 XX
 XX The present invention describes a polynucleotide (I) comprising a
 CC sequence of a Bifidobacterium genome selected from the nucleotide
 CC sequences given in AB081842 and AB081843, or a sequence exhibiting at

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2005, 23:47:50 ; Search time 105 Seconds

(Without alignments)
2026.661 Million cell updates/sec

Title: US-09-750-986D-30

Perfect score: 3082
Sequence: 1 MRLEALPRGRILRLH.....KVDALYRGDSQMLRDEATL 589Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/prodata/2/pubppaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/prodata/2/pubppaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/prodata/2/pubppaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/prodata/2/pubppaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/prodata/2/pubppaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/prodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/prodata/2/pubppaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/prodata/2/pubppaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/prodata/2/pubppaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/prodata/2/pubppaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/prodata/2/pubppaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/prodata/2/pubppaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/prodata/2/pubppaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/prodata/2/pubppaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/prodata/2/pubppaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/prodata/2/pubppaa/US10D_PUBCOMB.pep:*
- 17: /cgn2_6/prodata/2/pubppaa/US10E_NEW_PUB.pep:*
- 18: /cgn2_6/prodata/2/pubppaa/US11_NEW_PUB.pep:*
- 19: /cgn2_6/prodata/2/pubppaa/US60_NEW_PUB.pep:*
- 20: /cgn2_6/prodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	3082	100.0	589 8 US-08-976-063C-30	Sequence 30, Appl
2	3082	100.0	589 10 US-09-750-986D-30	Sequence 20490, A
3	1457	47.3	575 14 US-10-369-493-20490	Sequence 11697, A
4	1294	42.0	576 14 US-10-369-493-11697	Sequence 14740, A
5	1294	42.0	576 14 US-10-369-493-14740	Sequence 15216, A
6	1294	42.0	576 14 US-10-369-493-15216	Sequence 14320, A
7	1250	40.6	553 14 US-10-369-493-14320	Sequence 17047, A
8	972	31.5	621 14 US-10-369-493-17047	Sequence 244, App
9	383.5	12.4	544 14 US-10-369-493-244	Sequence 13094, A
10	333	10.8	636 14 US-10-156-761-13094	Sequence 10386, A
11	324	10.5	582 14 US-10-369-493-10386	Sequence 17, Appl
12	283	9.2	700 10 US-09-906-419-17	Sequence 130, App
13	283	9.2	700 14 US-10-119-136-130	

14	283	9.2	700 15 US-10-410-031-130	Sequence 130, App
15	281	9.1	585 14 US-10-369-493-311	Sequence 311, App
16	281	9.1	613 14 US-10-369-493-19735	Sequence 19735, A
17	279	9.1	700 14 US-10-119-136-17	Sequence 17, Appl
18	279	9.1	700 15 US-10-410-031-17	Sequence 17, Appl
19	278.5	9.0	613 14 US-10-369-493-20886	Sequence 20886, A
20	278	9.0	562 14 US-10-369-493-13736	Sequence 13736, A
21	278	9.0	701 10 US-09-906-419-16	Sequence 16, Appl
22	278	9.0	701 14 US-10-119-136-16	Sequence 16, Appl
23	278	9.0	701 15 US-10-410-031-16	Sequence 16, Appl
24	277.5	9.0	504 14 US-10-369-493-12089	Sequence 12089, A
25	277.5	9.0	598 14 US-10-369-493-8109	Sequence 8109, App
26	273	8.9	677 17 US-10-739-930-6854	Sequence 6854, App
27	272	8.8	589 14 US-10-369-493-21245	Sequence 21245, A
28	268	8.7	489 14 US-10-369-493-20547	Sequence 20547, A
29	268	8.7	500 14 US-10-369-493-19511	Sequence 19511, A
30	265	8.6	601 14 US-10-369-493-923	Sequence 923, App
31	265	8.6	638 14 US-10-156-761-12352	Sequence 12352, A
32	265	8.6	6842 15 US-10-461-194-131	Sequence 131, App
33	264.5	8.6	546 10 US-09-838-469-23	Sequence 23, Appl
34	263.5	8.5	544 14 US-10-378-168-23	Sequence 23, Appl
35	263	8.5	634 15 US-10-713-810-3	Sequence 3, Appl
36	263	8.5	696 17 US-10-425-115-300648	Sequence 300648, Sequence 165496,
37	263	8.5	858 16 US-10-437-963-165496	Sequence 24, Appl
38	262.5	8.5	544 10 US-09-838-469-24	Sequence 2, Appl
39	262.5	8.5	544 10 US-09-813-2798-4	Sequence 24, Appl
40	262.5	8.5	544 14 US-10-378-168-24	Sequence 45, Appl
41	262.5	8.5	544 14 US-10-378-168-45	Sequence 2, Appl
42	262.5	8.5	544 16 US-10-655-878-2	Sequence 4, Appl
43	262.5	8.5	544 16 US-10-655-878-4	Sequence 67862, A
44	262.5	8.5	4317 15 US-10-282-122A-67862	

ALIGNMENTS

RESULT 1
US-08-976-063C-30
Sequence 30, Application US/08976063C
Publication No. US20020182697A1
GENERAL INFORMATION:
APPLICANT: Alexander Steinhuebel; Horst Priefert; Jurgen Rabenhorst
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
TITLE OF INVENTION: CONFERTIL ALCOHOL, CONFERTILALDEHYDE, FERULIC ACID, VANILLIN AL
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: HP VECTRA
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976, 063C
FILING DATE: 21-NOV-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 196 49 655.1 (Germany)
FILING DATE: 29-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844

TELEX:
 ; INFORMATION FOR SEQ ID NO: 30:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 589 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-976-063C-30

Query Match 100.0%; Score 3082; DB 8; Length 589;
 Best Local Similarity 100.0%; Pred. No. 7.7e-289;
 Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MRSLEALPPPGRIERLEHMAKTRPEQTCVAAARAANGEMRRI SYAEFHNVAIAQSLL 60
DB 1 MRSLEALPPPGRIERLEHMAKTRPEQTCVAAARAANGEMRRI SYAEFHNVAIAQSLL 60
QY 61 PYGLSARPLLIVSGNDLEHLQAFGAMYAGIPYCVSPAYSLISODLAKRHIVGLQP 120
DB 61 PYGLSARPLLIVSGNDLEHLQAFGAMYAGIPYCVSPAYSLISODLAKRHIVGLQP 120
QY 121 GLVFAADAAFPORAITETIIPDDVPALFTRGELAGRTVSFDSLLEPGGIEADNAFAATG 180
DB 121 GLVFAADAAFPORAITETIIPDDVPALFTRGELAGRTVSFDSLLEPGGIEADNAFAATG 180
QY 181 PDTIAKFLFTSGSTKLPAKVPPTQRMLCANQOMLQTFPVFGSEPPVLDMLPMNHTFGG 240
DB 181 PDTIAKFLFTSGSTKLPAKVPPTQRMLCANQOMLQTFPVFGSEPPVLDMLPMNHTFGG 240
QY 241 SHNIGIVLYNGGTYLDDGKPTAQGFATLRNLSEISPTAYLTVPKGMEELVGALERDST 300
DB 241 SHNIGIVLYNGGTYLDDGKPTAQGFATLRNLSEISPTAYLTVPKGMEELVGALERDST 300
QY 301 LRERFPAARMKLFPPAAAGISOGIMDRLDVABEOHCGERIRMMAGLQMTETAPSCFTTGP 360
DB 301 LRERFPAARMKLFPPAAAGISOGIMDRLDVABEOHCGERIRMMAGLQMTETAPSCFTTGP 360
QY 361 LSWAGYIGLPAPGCEVKLPVVDGKLEGRFPHGPHVMSGYRPAPEONQAFDEEGYCSGDA 420
DB 361 LSWAGYIGLPAPGCEVKLPVVDGKLEGRFPHGPHVMSGYRPAPEONQAFDEEGYCSGDA 420
QY 421 IKLADPADPOKGMFDRGRIAEDEFKLSGVFVSGPLRTAVALLEGGSYLDVVVAADPRRC 480
DB 421 IKLADPADPOKGMFDRGRIAEDEFKLSGVFVSGPLRTAVALLEGGSYLDVVVAADPRRC 480
QY 481 LGLVLPRLDLCRALSGLGEKASDAEVLASEPVRAMPADMLKRLNREATGNASRIWVGL 540
DB 481 LGLVLPRLDLCRALSGLGEKASDAEVLASEPVRAMPADMLKRLNREATGNASRIWVGL 540
QY 541 LDTPPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDQSMLRDEATL 589
DB 541 LDTPPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDQSMLRDEATL 589

```

RESULT 2

US-09-750-986D-30
 ; Sequence 30, Application US/09750986D
 ; Publication No. US20030228670A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Steinbuechel, Alexander
 ; APPLICANT: Pfelefer, Horst
 ; APPLICANT: Rebenhorst, Jurgen
 ; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONFIERUL
 ; TITLE OF INVENTION: ALCOHOL, CONFIERULDEHYDE, FERULIC ACID, VANILLIN AND
 ; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
 ; FILE REFERENCE: Bayer-9998-CAO
 ; CURRENT APPLICATION NUMBER: US/09/750, 986D
 ; PRIOR FILING DATE: 2000-12-28
 ; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 30

LENGTH: 589
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas sp.
 ; US-09-750-986D-30

Query Match 100.0%; Score 3082; DB 10; Length 589;
 Best Local Similarity 100.0%; Pred. No. 7.7e-289;
 Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MRSLEALPPPGRIERLEHMAKTRPEQTCVAAARAANGEMRRI SYAEFHNVAIAQSLL 60
DB 1 MRSLEALPPPGRIERLEHMAKTRPEQTCVAAARAANGEMRRI SYAEFHNVAIAQSLL 60
QY 61 PYGLSARPLLIVSGNDLEHLQAFGAMYAGIPYCVSPAYSLISODLAKRHIVGLQP 120
DB 61 PYGLSARPLLIVSGNDLEHLQAFGAMYAGIPYCVSPAYSLISODLAKRHIVGLQP 120
QY 121 GLVFAADAAFPORAITETIIPDDVPALFTRGELAGRTVSFDSLLEPGGIEADNAFAATG 180
DB 121 GLVFAADAAFPORAITETIIPDDVPALFTRGELAGRTVSFDSLLEPGGIEADNAFAATG 180
QY 181 PDTIAKFLFTSGSTKLPAKVPPTQRMLCANQOMLQTFPVFGSEPPVLDMLPMNHTFGG 240
DB 181 PDTIAKFLFTSGSTKLPAKVPPTQRMLCANQOMLQTFPVFGSEPPVLDMLPMNHTFGG 240
QY 241 SHNIGIVLYNGGTYLDDGKPTAQGFATLRNLSEISPTAYLTVPKGMEELVGALERDST 300
DB 241 SHNIGIVLYNGGTYLDDGKPTAQGFATLRNLSEISPTAYLTVPKGMEELVGALERDST 300
QY 301 LRERFPAARMKLFPPAAAGISOGIMDRLDVABEOHCGERIRMMAGLQMTETAPSCFTTGP 360
DB 301 LRERFPAARMKLFPPAAAGISOGIMDRLDVABEOHCGERIRMMAGLQMTETAPSCFTTGP 360
QY 361 LSWAGYIGLPAPGCEVKLPVVDGKLEGRFPHGPHVMSGYRPAPEONQAFDEEGYCSGDA 420
DB 361 LSWAGYIGLPAPGCEVKLPVVDGKLEGRFPHGPHVMSGYRPAPEONQAFDEEGYCSGDA 420
QY 421 IKLADPADPOKGMFDRGRIAEDEFKLSGVFVSGPLRTAVALLEGGSYLDVVVAADPRRC 480
DB 421 IKLADPADPOKGMFDRGRIAEDEFKLSGVFVSGPLRTAVALLEGGSYLDVVVAADPRRC 480
QY 481 LGLVLPRLDLCRALSGLGEKASDAEVLASEPVRAMPADMLKRLNREATGNASRIWVGL 540
DB 481 LGLVLPRLDLCRALSGLGEKASDAEVLASEPVRAMPADMLKRLNREATGNASRIWVGL 540
QY 541 LDTPPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDQSMLRDEATL 589
DB 541 LDTPPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDQSMLRDEATL 589

```

RESULT 3

US-10-369-493-20490
 ; Sequence 20490, Application US/10369493
 ; Publication No. US2003023675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 20490
 ; LENGTH: 575
 ; TYPE: PRT
 ; ORGANISM: Rhodospseudomonas palustris
 ; US-10-369-493-20490

Query March	47.3%;	Score 1457;	DB 14;	Length 575;
Best Local Similarity	50.5%;	Pred. No. 1.5e-131;		
Matches 292;	Conservative 86;	Mismatches 194;	Indels 6;	Gaps 2;

```

QY      1 MRLEALPPGGILRLREHMAKTREQTCVAARANGCWRRISVEMFHNRAAQSLI    60
Db      1 VRSTTTLTVPVATIDRLNHFATTADRYFMAMERNEGGRNISYLEMURAQTASALI   60
QY      61 PYGLSAPERLLIVSGNDLHLQLAFAMTAGTIPYCSPAYSLSLODLAKLRHVIGLIOP  120
Db      61 ARGLSAERPVMIIISGNSIDHAMVFGLAYAGVMCFVSPPYSIVSKDYKCLSHIIGLLTP  120
QY      121 GLVFAADAAPFGQAIFETLPDDVRPAFTTGELAGRRTVPSDSLIEPGGI EADNAFPATG  180
Db      121 GLLPADDITAFAPAILLATVPEDELATATGEVGKRVTLSLAIELATTEHPRELAAKEHGAI  180
QY      181 PDTIAKFLPTSGSTKLPKAVPTTORMLCANQMILLQTFPVFGBEPPVLVDMLPMNHTEGG  240
Db      181 HDTIATAKFLISSGSTGNMPKAVINIQRMICANQWVIREAMFLNDEPPVIVIDMLPMNHTEGG  240
QY      241 SHNIIGIVLVNGGYTYLDDCKPTAQGAERTIRNLSEISPAIYLTVPRGMVELVGLERDST  300
Db      241 NNHIIGILTFFNGSGMYIDDGKPRTAGIASITRINLREIAPVFNVPVKGSLLPLVIREDOQ  300
QY      301 LRERPFARMKLRFEPAAAGLSQGIMDRLDRYAEQHCEERI RMAGLGMTETA PSTCTFTGP  360
Db      301 LRLTLFSTRLLHAMPFSQASLAAHVNWGLDEVAYALETARKPMLTG LGATITTA PFMSVTPQ  360
QY      361 LSMAGYTGLPARCCEVKLYPVNDGLEGRFHGHVMSGVWRAPEDQNAAFDEEGYCSGDA  420
Db      361 TSSRGHWGLPEVPENENKLVPNNGKLEVARAKGPIITGVYRAPBLDTKADEEGRFYKLNDA  420
QY      421 IXLADPADPOXGLMFQRGTAEDEFKSSGVSVSGPLRTAVJLEGSSYYLVLDVVVAADPREC  480
Db      421 LRPVDANDUSRGDFPDRI SEDFKLASGTWVSVPRLPAKFIAACSLAVDVVIAGLDRXY  480
QY      481 LGLVLFPRILD--CRALSGLGKEASDAEVLA SEPRAMFADMCLKLNREATGNARSIRMW  537
Db      481 VTALA---LIDDPGCULINATLPLEDLAGAADHLIREAFREFRFA TLTLQATGSSNRVTR  537
QY      538 VGLDTTPPSIDKGEBVTDKGSI NORAVYLQMRSAKYDALY  575
Db      538 AVLIGEPLSIDKGEITDKGSVNORAVLEYRASLIADLY  575

RESULT 4
US-10-369-493-11697
| Sequence 11697, Application US/10369493
| Publication No. US20030233675A1
| GENERAL INFORMATION:
| APPLICANT: Cao, Yongwei
| APPLICANT: Hinkle, Gregory J.
| APPLICANT: Slater, Steven C.
| APPLICANT: Goldman, Barry S.
| APPLICANT: Chen, Xianfeng
| TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
| FILE REFERENCE: 38-10(52052)B
| CURRENT APPLICATION NUMBER: US/10/369, 493
| PRIOR FILING DATE: 2003-02-28
| PRIOR APPLICATION NUMBER: US 60/360, 039
| NUMBER OF SEQ ID NOS: 47374
| SEO ID NO 11697
| LENGTH: 576
| TYPE: PRT
| ORGANISM: Agrobacterium tumefaciens
| US-10-369-493-11697
```

[illegible]

```

Oy      5 EALLPEPPGRIRLERLEHWAKTREBOTCVARAANGEMWRRISYAEMHNVRALAQSLIPYGL 64
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1 DPLGPPYQKUNLRRLLHWCRSADERTWMDRQGRPPWRRRSYAEALDKIRIRIGQFLDDHDL 60

```

QY 65 LAERELLVSGNDLQLQAFGMVAGIPYCPSPVSYLSLOLAKRHHVGLLOPGIVF 124
 Db 61 SVBRDLVLSSENSIEIALMVLAAQHVIGISAATIPAA-TSADITTKADIRGQITPGKVF 119
 QY 125 AADAPFORAIEITLIPDDVPAIFTRGELAGR-RTVSPDSLLEOPGGIEADNAFAATGPD 183
 Db 120 AEDATPFRRALGEVDPDGRPLVGLRNLPEDRNSTPFHETLLETTEPTAVDRAFPDAPDT 179
 QY 184 IAKFLFTSGSTKLPRKAVPTTORMLCANQOMLLOTFPVYGEERPYLDVLMFNHTTGGSHN 243
 Db 180 VAKPLFTSGTTGSPYAAVLOTORMLCNNOEMLADOCYGFREEPVVDMAWPNHTAAGNKV 239
 QY 244 IGIVLVNGSTYVLLDDCKPTAAGFAELRLNLSISPTAYLTVPKGMEELVGLAREDSLRE 303
 Db 240 FNLVLVNGSTYVLLIDRKEPPAQTGOTLDNRDLSPTMYENVDPGHEMLVQMRKDEALCR 299
 QY 304 RFFAAMKLFPPAAGLSQGIWDLRLDRAVBOHCCEBRLRMAGLMTETAPSCFTTGPLSM 363
 Db 300 SFPRDLKMLVAGAGMAQHTWDALTELMAVTAHVAVLGMAGDLSSTERAPSLFTBPDQK 359
 QY 364 AGYILPAPGCYVLLPVNDKLEGRFHGPHVMSGYRARPQNAQAFBEGYCSGDAIKL 423
 Db 360 PGNIGIPAGQYTMKLVPPFDGRYELRLUKGPVITPGYRNNGELTAAAFDEBEPYRIGDVTKF 419
 QY 424 ADPADPOGLMADGRIAEDPFKLSGGVFSVGPRTPAVLEGGSVLDDVVVAAPDREGGL 483
 Db 420 AVADDPKRGFYDGRKAEVFKLOTGTWAVAGPLRAQVLMNFAGLIRDAVLTGGENRAELGA 479
 QY 484 LVFPRLLDORALSGLKEASDAEVLASEPYRAMFADWLKRLNEATGNASRIMVWGLLDT 543
 Db 480 LVFPRIPALRELVRSGSHLSDAEIIHHPVRAQIVAKLSHQASGSASRVMRLVWED 539
 QY 544 PPSIDKEVTDKGSINQRAVLDQMSKADVALY 575
 Db 540 ALRFEKGEVTDKGSINQRAVLLHRKELVBSLY 571
 RESULT 6
 US-10-369-493-15216
 / Sequence 15216, Application US/10369493
 / Publication No. US20030233675A1
 / GENERAL INFORMATION:
 / APPLICANT: Cao, Yongwei
 / APPLICANT: Hinkle, Gregory J.
 / APPLICANT: Slater, Steven C.
 / APPLICANT: Goldman, Barry S.
 / APPLICANT: Chen, Xianfeng
 / TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 / FILE REFERENCE: 38-10(52052)B
 / CURRENT APPLICATION NUMBER: US/10/369, 493
 / CURRENT FILING DATE: 2003-02-28
 / PRIOR APPLICATION NUMBER: US 60/360, 039
 / PRIOR FILING DATE: 2002-02-21
 / NUMBER OF SEQ ID NOS: 47374
 / SEQ ID NO 15216
 / LENGTH: 576
 / TYPE: PRT
 / ORGANISM: Agrobacterium tumefaciens
 / US-10-369-493-15216

Oy	125	AAAPAAQRAIETLLDDVPAIFETRELAGR-RTVSPSLIEQPGCIEADNAFAATGPD	183
Db	120	AEDTTPRRRLGAVFPDGTPLVGLRNLPRDRSNTTFETLLETETLEAVDRAFDVAGPDT	179
Oy	184	IAKFLFTSGSTKLPKAVPTTORMLCANQOMLLQTEPVEFGEEPPVLVDLPMNHTFGSSHN	243
Db	180	VAKFLFTSGTGSPPKAVIQORMLCNQEIMADCGYFEEEPVVVDNAPMHTAAGNKV	239
Oy	244	IGVLVNGGTYVLDGDKPTAQAQPAETRLSLISPAVLYTVKGMELVGLAERDSTLRE	303
Db	240	FNVLVNLGGTYVLDKRPSPAQIGQLTLRLDISPMWFFNVPAQHMLVQARKKDALCR	299
Oy	304	RFEARMKLFFPFAAAGLSOGIWDRLDRVAEOHCGERIIMNAGLMTETA PSTCTTGPLSM	363
Db	300	SFFRDLKMLMVAAGAAQHTWDLTSLMATVGHAVLMGAGLSTETAPSLFCTEPDQK	359
Oy	364	AGYIGLPAPCECYLVPVNGKLEGRPHGHVMSGYTRAPEDQNAOPDEEGYCSGDATKL	423
Db	360	PGNIGIPAOQVTKMLVPFDGRYELRKGNITIPGWYRNKELTPAAPDEEGFRIDITVKP	419
Oy	424	ADPADPOKGMFDPGIAEDDFKLSGQVGVSVGLRTAVALLEGSSVYLDVVAAPDRECLGL	483
Db	420	AVADDPRRGVYFPGRAENPKLQGTQWVAVGLRLAQVLMMFAGLLIRDAVITGENNAELGA	479
Oy	484	LVFPRLLDCALSGLKEASDAEVLASEPVRAWFADWLKLNREKATGNASRLMMVGLLDT	543
Db	480	LVVFPIPALNELVRGSHLSDAEIINHPSVRAQIVAKLSAHQKASGSASRYMLIVMED	539
Oy	544	PSIDKGEVTDKGSINQRAVLCWRSKAVDALY 575	
Db	540	ALRFKEQEVTDKGSINQRAVLLHREKELVSLY 571	
RESULT 7			
US-10-369-493-14320			
Sequence 14320, Application US//10369493			
Publication No. US20030233675A1			
GENERAL INFORMATION:			
APPLICANT: Cao, Yongwei			
APPLICANT: Hinkle, Gregory J.			
APPLICANT: Slater, Steven C.			
APPLICANT: Goldman, Barry S.			
APPLICANT: Chen, Xianfeng			
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF			
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES			
FILE REFERENCE: 38-10(52052)B			
CURRENT APPLICATION NUMBER: US/10/369,493			
PRIOR FILING DATE: 2003-02-28			
PRIOR FILING DATE: 2002-02-21			
NUMBER OF SEQ ID NOS: 47374			
SEQ ID NO 14320			
LENGTH: 553			
TYPE: PRT			
ORGANISM: Agrobacterium tumefaciens			
US-10-369-493-14320			

Query Match	42.0%	Score 1294	DB 14	Length 576
Best Local Similarity	46.5%	Pred. NO 9e-116		
Matches 266	Conservative 84	Mismatches 220	Indels 2	Gaps 2

QY	5	EALLPFCRILIERLEHNAKTRPEQTCVAPARANGEMRIRISAEHMFHNVRAIAQSLIPYL	64
DB	1	DLPLDYPQKLNERRLLHWCGRSAPEIRMMADRGREBWRVSYAEALDKTRRQGFLLDDDL	60
QY	65	SAERPLLIVSGNDDEHLQALFAGANTAGIPYCPVSPAYSLSLQODAKLTHIYGLLQPLGVF	124
DB	61	SVREPLPLVLSNENSLHEHMLMVAAGHVGASAAIRPAYA-TGADITLKADINGQITPGGVF	119

Query Match	Similarity	40.6%	Score 1250;	DB 14;	Length 553;
Best Local	Similarity	46.2%	Pred. No. 1.5e-111;		
Matches	Conservative	82;	Mismatches 214;	Indels	Gaps 2;
Qy	5	EALLPFGRIILRIEHWAKTRPEQTCVAARANGEMWRISVAYEMFHNVRATAGSLPGL	64		
		: : : : : : : : : : : : : : : : : : : : : : : : :			
Db	1	DPLOPYQKLNRIELHHCRSAPERTWADRGRRPWRVSYAEALDKIRRGQFLDHDL	60		
Qy	65	SAERPLIVSGNDLEHLQLAFGAMYAYIPYCPVSAYSLSQDLAKLRHIVGLQGLVLF	124		
		: : : : : : : : : : : : : : : : : : : : : : : : :			
Db	61	SVERPLIVLSNSISLHMLVLAQHGVIASAIRIPAYA-TSADLTKLADIRGQITPGWVF	119		
Qy	125	AADAPAPORAIETTLPPDVPAIFPRGELAG-RIVSFDLSLLEOGGIELADNAPATGDT	183		
		: : : : : : : : : : : : : : : : : : : : : : : : :			
Db	120	AEDTTPRRALAGEVFDGTPLVGLRNLPEDRSNTEFHELTLETETPEAVDAPFAVGDT	179		
		: : : : : : : : : : : : : : : : : : : : : : : : :			

[illegible]

QY 315 AAAGISQGIWRLDRAVDHCGERIRBMAGLMTETAPRCTFTTGGLSMAGY-----IGL 369
Db 304 AASKLPPIEDYF-----KHCG--IYVABGKGTETSP-CLTIKGP-----GYEHNVSGR 351
QY 370 PAPCGEVLVPVYDGLKLGRRFHGPHWMSGYRARPEONAAFDDEGYGCSGDATKIADPADP 429
Db 352 PLPNVYLIDSDSGEILAK--GPNWMLGYADEAANQRCF-RQGWERTDGLKMI-----403
QY 430 OKG-LMPFGRIADPEKLSGGSFVSGPLFTRAVLEGGSVYLDVWVAAPRECGILVF--486
Db 404 -KGELVILGRLDSYKLSGKEVSSDALE-QSLFEKTDIIINFAIIDIEQRPATALLFIN 461
QY 487 PLLDLCRALSGLKEASDAEVLASEBVRAMFMDKRLRLREATGNASRIIMWYGLDTPPS 546
Db 462 PEL--CRQRFG-----ENCTLGHSRLGHIAIVAVQDHNQYEDSSFKLIAVAVETSTLS 514
QY 547 IDKGEVTPDKSGINQRAVYLQMSAKYDALY 575
Db 515 LEKGEVTPSFVKQSQRTVRENNQYADVALY 543

```

RESULT 10
US-10-156-761-13094
; Sequence 13094, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13094
; LENGTH: 636
; TYPE: PR1
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-13094

```

```

Query Match      10.8%; Score 333; DB 14; Length 636;
Best Local Similarity 23.7%; Pred. No. 1.1e-22;
Matches 151; Conservative 100; Mismatches 261; Indels 124; Gaps 25

QY      PEQTCVAAARAANGEMRRISYAEFMFNVRVIAQSLPLPYGLSAERPLIVSGNDLEHLQALAF 85
      |||      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      PEGGAPRAGQGPDDMKSLSMQAARVRVPAVAGLITLGVQPEGRVALASTREWILLADL 104
      45 PEGGAPRAGQGPDDMKSLSMQAARVRVPAVAGLITLGVQPEGRVALASTREWILLADL 104

QY      GAMAAGIPYCPVSPAYSLLSODLAKRHIVGLLQGPVLEPADAPFORAIE----- 136
      |||      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      GIMCAGAAATTTIYF-----QTNVADSAPILADSESKVLLAEPDAQLAKKVERKAELPAL 159
      105 GIMCAGAAATTTIYF-----QTNVADSAPILADSESKVLLAEPDAQLAKKVERKAELPAL 159

QY      ---TLPPDVPA---IFTRGELAGRRYTSF---DSLRLQPGGIEADNAPFAATGPDTIAK 186
      |||      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      HVVVVDIPFGVESSEDMVLTLLAELEARGAAHLKKNPDLIKKKVGAITSDQ-----LAT 210
      160 HVVVVDIPFGVESSEDMVLTLLAELEARGAAHLKKNPDLIKKKVGAITSDQ-----LAT 210

QY      FLFTSGSRKLPKAVPTTQRMLCANQOMLLQTFPRVG--EEPRVLVWMLPNNHTFGGSHNI 244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      LIYTSGITGRPKGV---RLPHDNNSYAKATAGTLLISADVDQYUMLPLAHFEGK---- 262
      211 LIYTSGITGRPKGV---RLPHDNNSYAKATAGTLLISADVDQYUMLPLAHFEGK---- 262

QY      GIVLYNGGTYYADDDKPTAQG--PAFTLNLSEISPTAVLYVKKGEELVGALEKSDTLR 302
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      --VLTSG--QIEVGHVTAVDGRVDVDTIENLPVQDTYMAAVPRITEKYNGA--AKAR 315
      263 --VLTSG--QIEVGHVTAVDGRVDVDTIENLPVQDTYMAAVPRITEKYNGA--AKAR 315

```

```

Oy      30  ERFPAFKLFFFA-----AGLSGIMDRD-----RVAEQ-----333
           ||: ||: ||:
Db      316 AGGAKYKILFMAAEVAREYAKVSGDNFRFRGAASVPFGLGAKKXVADALVFAKIREAFG 375
Oy      334  -----HCGERIRMMAGLGMETPASTCTFTTGPLSMAGYILGPARG 373
           ||: ||: ||:
Db      376 GNLRAVCSSAALAPBEGYFFPAGAGIHLEGGLTSSSAPSVNPGAYRTGVGKPLPG 435
Oy      374  CEYKLVAVPGKLEGRPHGPHVMVSGWAPAEONACAFDEEGYCSGDAIKLADPADPQGL 433
           ||: ||: ||:
Db      436 TEVRIND-DGET--LLRGGIMEGHGRLPXTAETLEBDEGMFHNGDIGELS----PGYL 488
Oy      434  MFDGRIAEDFKLSSGFVUSVGPLRTRAVLEG-GSYLVDVVAADRE-CIGLVF--PRL 489
           ||: ||: ||:
Db      489  RITDRKKDILIKTSGGYIA--PAEVEGQFKGCVPSVSNIVHGDNRNFTALLALDEPSI 546
Oy      490  LDCRALSGP-GKEASDAEVLASEPRAPFADWLKRLNEATGNASR--IMWGLDTPP 545
           ||: ||: ||:
Db      547  LEWAKENGLAGK--SYADVVAAPATVALVEGVTELN----GGLQRMQTIKKPRLPRDL 600
Oy      546  SLDKGEVTDKGSINCRAYLQMRSAKYVDLNYGEDOS 581
           ||: ||: ||:
Db      601  DIHGEITPSLAKRPVEREYKHLIDIMWYGSRA 636

```

```

RESULT 11
US-10-369-493-10386
; Sequence 10386, Application US/10369493
; Publication NO. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10386
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Cytophaga hutchinsonii
US-10-369-493-10386

```

Query Match	10.5%	Score 324	DB 14	Length 582
Best Local Similarity	23.6%	Pred. No. 7.2e-22		
Matches 144	Conservative 101	Mismatches 264	Indels 100	Gaps 26
QY	26	PEQCYAABAANGEMWRISYAEMFNVAIAQSLPYGLSAERPLLIVSGNDLEHLQIAR	85	
DB	13	PKQDIALATK-ENGCTWISYSVAVQEIIDKVGISGIFKSLQDKDVKVAILSPRPENMIDL	71	
QY	86	GANYAGIPYCPVSPAYSLSDQDLAKRHIVGSLQGLVFAADAAPFORAIEITIIIPDV-P	144	
DB	72	GLQQIGAVSVPIYP-TLIVED--YRIYIPDAEAVKVFYVDEQLFAKATEAVNGSRNP	126	
QY	145	AITRREGELAGRTVSFDSLLE--QGGGIEADNAPFAICPDIIAKFLFTSGSTKLKPAVPT	202	
DB	127	VITLLDQVAVGASWKDITLADGGPASM-LDPYKANVFNDLTLITYSGTGPKGVML	185	
QY	203	TQRMLCANQOMLQTFPVPGSEEPPLVDMLPMNHTFGGSHHIGLVY--NGGTYYLDDGK	260	
DB	186	THNNILSNYIAAIPLMPV--NNNHKALSFLPLCHVY--ERMCLCYLLSGQSVSIYVAEST	240	
QY	261	PTAQGFPAETLRNLSEISPTAYLIVTEPKGE-----ELVG-----ALER	297	
DB	241	DTVG-----ENIKVEQPMFVTVVRLLIEKYDKIVAKGAEITGIKKFLPFWALEGLKY	294	

KM Ligate. 632 AA; 68280 MW; C082A41146934F7 CRC64;
SQ SEQUENCE

Query Match 47.3%; Score 1457; DB 2; Length 632;
Best Local Similarity 50.5%; Pred. No. 1.1e-97;
Matches 292; Conservative 86; Mismatches 194; Indels 6; Gaps 2;

1 MRSLEALLPPGGILRLLEHMAKTRPEQTCVAAARAANGERRISYAEFMFNVAIAQSLL 60
46 VNSTTTLTDPVITTDRLHFAETAPRVMAERNGCGRRRSYAEMLRAAQITIASALI 105
61 PYGLSARPLLIYSGNDLEHLQAFGAMYAGIPYCPVSPAYSLSODLAKRHIVGLIOP 120
106 ARGLSARPMILISGNSIDHAMMFGALYAGVAMCPVSPYLSVKDYGKLRHIVGLITP 165
121 GLVFAADAAFPQRAIETILPDDVPATFTRGELAGRTVSPDSLBOGIEADNAFAATG 180
166 GLIFADDTAFAPAIATVPEDEVELAATRGVGRKVTSLAEILATPEHPELAAKHEAIG 225
181 PDITAKPLFTSGSTKLPKAVPTTQRMICANQOMLQTFPVGGEPPVLVDMLPNNHTFGG 240
226 HDITIAFLTLTSGSTGNPKAVINTORMICANQVMTREMAFLKDEPPVITDMLPNNHTFGG 285
241 SHNIGIVLYNGGTYLLDCKPTAGFAETLRNLSEISPTAYLTVPKGMEEVLGALERTST 300
286 NNHIGLTFNGSGMYIDGKPTAGIASTIRNLREIAPTYFNVPKGVESLPLVLRDQ 345
301 LRRFRFAKMLFFPAAAGLSQGIWDRIDRYAEOHGERIRMMAGLMTETAPSCFTTGP 360
346 LRKLFFSRLHMFSGASLAHAWNGIDEVAVAETGARVPMLTGLGATEPAPFMSVTPQ 405
361 LSMAAGTIGLPAPCEVVLVVDGKLEGRFHGPHVMSGYWRAPBQNAAPFDEEGYCSGA 420
406 TSSGHHGLVPGNEAKLVNNGKLEVRAGVPNTTPEYMPAPETIDYAFDEEGFYKLANDA 465
421 IKLADPADPQKGLMFDGRIAEDEFLSSGVFVSGVPLTRAVLEGGSYVLDVVVAAPDREC 480
466 LKVDANDLSRGFDPOGRISEDFLKAGTWSVGLPAKFLAACSLVRRVVIAGDRDY 525
481 LGLLVFPRLLD--CRALSGLEKASDAEVLASEPVRAMPADWIKLNRATGNASIRIMW 537
526 VYALA--LIDPDGCKIKNTLPLEDLAGNAADHLIREAFREPRATLTQATGSSNRVTR 582
538 VGLLDPPTSDKGEVTDKGSINORAVLQWRSAYVDALY 575
583 AVILGEPLSIDKGEITDKGSVQRAVLEVRASLIADLY 620

Db

RESULT 13
ID 089CF2 PRELIMINARY; PRT; 627 AA.
AC 089CF2;
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Long-chain-fatty-acid-CoA-ligase.
OS OrderedLocuNames=bl17845;
OC Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN RAIN
RP SEQUENCE FROM N.A.
RC STRAIN=USDA110;
RX MEDLINE=2248498; PubMed=12597275;
RA Kaneo T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriyuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
CC -I- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme

CC family.
DR EMBL: A005963; BAC53110.1; -;
DR GO: GO:0016874; F.ligase activity; IEA.
DR GO: GO:0008152; F.mecabolism; IEA.
DR InterPro: IPR000873; AMP-bind.
DR Pfam: PF00501; AMP-binding; 1.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00455; AMP BINDING; 1.
KM Complete proteome; Ligate.
SQ SEQUENCE 627 AA; 68213 MW; F8C8CE304D38F88 CRC64;

Query Match 45.5%; Score 1401; DB 2; Length 627;
Best Local Similarity 50.7%; Pred. No. 1.3e-93;
Matches 293; Conservative 80; Mismatches 199; Indels 6; Gaps 4;

1 MRSLEALLPPGGILRLLEHMAKTRPEQTCVAAARAANGERRISYAEFMFNVAIAQSLL 60
41 LRKQPLGQVSVITRLLRHMAWTTTDPVFMEREGRGKRTTYAELLTASHHIASALI 100
61 PYGLSARPLLIYSGNDLEHLQAFGAMYAGIPYCPVSPAYSLSODLAKRHIVGLIOP 120
101 QRLSADRPVILISGNSIDHALAFQAFYAGVFCFVSPAYSLSVKDYGKLSLMLITP 160
121 GLVFAADAAFPQRAI--ETILPDDVPATFTRGELAGRTVSPDSLBOGIEADNAFAA 178
161 GLVFAADADKFDALANVSLGTEIASY--GHVAGRDVTLADLMAATPIRGDLDEVHGK 218
179 TGPDTIAKPLFTSGSTKLPKAVPTTQRMICANQOMLQTFPVGGEPPVLVDMLPNNHTF 238
219 IGDITIAKPLFTSGSTGNPKAVINTORMICANQVMTREMAFLKDEPPVITDMLPNNHTF 278
239 GGSNIGIVLYNGGTYLLDCKPTAGFAETLRNLSEISPTAYLTVPKGMEEVLGALERTST 298
279 GGSNHHGLTNGSGMYIDGKPTAGIASTIRNLREIAPTYFNVPKGVESLPLVLRDQ 338
299 STLRERFFARMKLFFPAAAGLSQGIWDRIDRYAEOHGERIRMMAGLMTETAPSCFTT 358
339 QGLRAKFPBRLHMFSGAALSPFVWDSLDELAVKEKGRVPMPLTGLGATEPAPFMSV 398
359 GPLSMAGTIGLPAPCEVVLVVDGKLEGRFHGPHVMSGYWRAPBQNAAPFDEEGYCSG 418
399 PRTSRGHGLVPGNDKALVPNNGLKLEVRAGVPNTTPEYMPAPETIDYAFDEEGFYK 458
419 DAIKLADPADPQKGLMFDGRIAEDEFLSSGVFVSGVPLTRAVLEGGSYVLDVVVAAPD 478
459 DALKPADPDDLANGPFDGVSDFLKAGTWSVGLPAKFLAACSLVRRVVIAGDRDY 518
479 BCLGLVFPRLD--CRALSGLEKASDAEVLASEP--VRAMPADWIKLNRATGNASIRIMW 537
519 DEVSALVVLDDGCRVLPNT--LEADDLVVATRDRLVREAFREPRATLTQATGSSNRVTR 577
538 VGLLDPPTSDKGEVTDKGSINORAVLQWRSAYVDALY 575
578 AILMDTPSIDKGEITDKGSVQRAVLEVRASLIADLY 615

Db

RESULT 14
ID 07CZ48 PRELIMINARY; PRT; 591 AA.
AC 07CZ48;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TEMBLrel. 27, Last annotation update)
DE AGR_C_2614P.
OS OrderedLocuNames=AGR_C_2614;
GN Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN RAIN
RP SEQUENCE FROM N.A.
RC STRAIN=Cereon;
RX MEDLINE=21608551; PubMed=11743194;

Qy	5	EALLPGRRIIELEHEMATRPEQOCVBARANGMRRIISVAEMHNRAIAOISLPLPG	64
Dh	35	EOLKAPFOKLTDBLIFPAKVCKPHTFFAAARNOTQDIEWIRISAEVTEIORMAHIAOALHORNL	94
Qy	65	SAERPLIVISGNLIEHLIOLAFGAMYAGIPEYCVSPAYSILLSODLAKLHIVIGLLQOPGVF	124
Dh	95	SGRPLVILISGNLIEHLITLSMGAMLAGVFSASISPAIVSIVQDFOKLGHVFDVLTPGMVY	154
Qy	125	AADAAFPORAIEITLPDDVPAIFTRGELAGRRTVSPDSILBOPGIEADNAPAAITGPDIT	184
Dh	155	ANNGOAFARXIOSCTODHIEIVTNKGIIDGYCTSFQSHLDP- VTDVQEHYEITLDEHOI	213
Qy	185	AKELFTSGSKLPEKAPPTQRMCANQOMLOTFPVPGEPPVLYDMYPMNHTGSGSHNI	244
Dh	214	AKELFTSGSKLPEKAPPTHTMLCINQOMLOTFPEFEDTPEPVLIDTMSMHTTGGSHNV	273
Qy	245	GIYLYNGATYLBDBGKPTAQGFATLRNLSEISPAVYLVTPGMEVLGALERDSTLRER	304
Dh	274	GIYLYNGATYIDDGKPEPKFDETRIRLKEISFPYLVANVPEGMELIDALEKALRDR	333
Qy	305	FFARMKLFPPAAAGLSQGIWDRIDRAVEOHCSERIIMNAGLMTETAPSCFTTGGPLSMA	364
Dh	334	FFANVKILFFAGALISEAGMNNRLDKIAOHCSEKIRISGLMTETAPSCAFTTGPVMA	393
Qy	365	GIYGLPAPGCVLVVVDGKLEGRFHFGRVMSGYMR-ABEQAOAFDEEGYTCGDAIKL	422
Dh	394	GIYGLPAPGCEIKLVVYGDKLEFCVRGKVMGYRLKADQOSTIFDEEGFFHGDVRL	453
Qy	424	ADPADPOKGLMPDRFAEDPKLSGFGVSVGLPRTFRAVLIEGGSYVLVVVAPAPRECGL	483
Dh	454	VDNDPAQGLMYGRFAEDPKLNTGTFVAVVGLTRKALKIIGNLLIODYCTIGSNLNAIGF	513
Qy	484	LVFPRLLDCRALSGIG-KEASDAEVLASVPVAMFADMLKRLNREATNAGRIIMVGLD	542
Dh	514	LVFPRLLDACADFAGLNMKHSALDILQHPKVQOMFROFLIYQNKDATGSSNRVSLYLMT	573
Qy	543	TPPSIDKEVYTDKGSINQPAVLOMSAKDALYREDOISM	583
Dh	574	EAPOLDAGEVTDKGNINQSNILKRRALVAEELVYNQOTNPL	614

RESULT 11			
06N8W6		PRELIMINARY;	PRT; 632 AA.
ID	Q6N8W6		
AC	Q6N8W6;		
DT	05-JUL-2004 (TRMBLrel. 27, Created)		
DT	05-JUL-2004 (TRMBLrel. 27, Last sequence update)		
DT	05-JUL-2004 (TRMBLrel. 27, Last annotation update)		
DE	Purative feruloyl-CoA synthetase (EC 6.2.1.-).		
GN	Name=fcsg2; OrderedLocustNames=RPAL787;		
OS	Rhodopseudomonas palustris;		
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;		
OC	Bradyrhizobiales; Rhodopseudomonas.		
OX	NCHI_TaxID=1076;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CGA009 / ATCC BAA-98;		
RC	PubMed=14704707; DOI=10.1038/nbt923;		
RA	Lamerin F.W., Chain P., Hauser L., Lamerdin J.E., Malfatti S., Do L.,		
RA	Land M.L., Pelletier D.A., Beatty J.T., Lang A.S., Tabita F.R.,		
RA	Gibson J.L., Gibson T.E., Bobb C., Torres y Torres J.L., Peres C.,		
RT	Harison F.H., Gibson J., Harwood C.S.;		
RT	"Complete genome sequence of the metabolically versatile		
RT	photosynthetic bacterium Rhodopseudomonas palustris";		
RL	Nat. Biotechnol. 22:55-61(2004).		
CC	-1. SMILIARITY: Belongs to the ATP-dependent AMP-binding enzyme		
CC	family.		
DR	EMBL; BX572598; CAB27228.1; -		
DR	GO; GO:0016874; F:ligase activity; IEA.		
DR	InterPro; IPR000873; AMP-bind.		
DR	Pfam; PF00501; AMP-binding; 1.		
DR	PRINTS; PR00154; AMPBINDING.		
DR	PROSITE; PS00455; AMP BINDING; 1.		

QV	SEQUENCE	632 AA;	68280 MW;	COB82A11469934F7 CRC64;
QV	Complete proteome; Lignase			
QV	Query Match	47.3%;	Score 1457;	DB 2;
QV	Best Local Similarity	50.5%;	Pred. No. 1,16-97;	
QV	Matches	292;	Conservative 86;	Mismatches 194;
QV			Indels 6;	Gaps 2;
QV	1	MRSLEALLPFPRGRLLERLEHNAKTRPPECVCYAAANAANGEMPRISYAEFHNVAIAOSLL	60	
QV	46	VRSTLTITLDYVVRTDRLHHPAEATAPRVFAENKNGEGWRRISYAEELRAQTIASALI	105	
QV	61	PYGSAERPLLIVSGNDLEHLQIAFGMYAGAIPLYCPSPAYSLISODLAKRHHVGLQOP	120	
QV	106	ARGISAERPVWILSGNSIDHAMVFGALYACVAMCPSPLSVSKDYGKLRHHVGLTTP	165	
QV	121	GLVEPADAAPFORAIEITLPDDVAIAFTRGELAGRTVSPFDSLLBQPGLEADNAFAATG	180	
QV	166	GLIFADDTITAPAPAIATVPEDEVELATRGVEVKRKTSLAELLATPEHPELAKHEAIG	225	
QV	181	PDTIAKELFTSGSKLKRKAVPTTORMCANOMLQTFPVVGEERPVLYMLPNNHTFGG	240	
QV	226	HDITAKELFTSGSGNPNPAVINTORMCANQVMIREMAFLKDEPVALVMDLPNNHTFGG	285	
QV	241	SHNIGIVLYNGGTYVLDGDKPTAOGFAETLNLSEISPTALVTPPKGEEELVGALEBDST	300	
QV	286	NHNIGITLTFNGSGMYIDDKPTPGIASTIRNLREINAPTYFNVPKGESLPLVIREQQ	345	
QV	301	LREPRFARMKLEFFPAAGLSQGIWDRLDRAVBQHCGERIRMMAGLGMETPASPCTFTTGP	360	
QV	346	LRLKLFESRLHAMPFSGASLAHVNNNGDEVAVETGARVPMLTGATETAPFEMSVPQQ	405	
QV	361	LSMAGVIGLAPRGEVULVPVDGLLEGRFHPHNSGWRAPRQNAQAFDEEGYVCSGDA	420	
QV	406	TSRSGHGVLPPVGAENKALVPPVNGKLEVRAPKNPTTPGWRAPELTDRAFDEEGYKLANDA	465	
QV	421	IKLADPADPOKGLAFNDRIAEDFSLSGGVFVSPLTRALTEGSGYVLDVVVAAPRECC	480	
QV	466	LKPPDANDLSNGPRPDRISDFLGAAGTWSVGPLAKFPIACASLVRODVIAIGLBDY	525	
QV	481	LGILLVPRLLD--CRALISGLKEASDAEVLASEPVAWPAWDLKRLNREATGNASRIW	537	
QV	526	VTALA---ILDPDCKILINATLPLEDLAGMAADHLIREAFERERFATLTLTQATGSSNRYTR	582	
QV	538	VGLDTPPSPIDKGEVTDKGSINQAVQMSAKYDALY	575	
QV	583	AVLLGGEPLSIDKGEITDKGSVNOQAVLEYRSLADIY	620	

```

RESULT 12
CAE27228 PRELIMINARY; PRT; 632 AA.
ID CAE27228
AC CAE27228;
DT 02-MAR-2004 (TEMBLrel. 27, Created)
DT 02-MAR-2004 (TEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TEMBLrel. 27, Last annotation update)
DE Putative feruloyl-CoA synthetase (EC 6.2.1.-).
GN PCS2 OR RPA187.
OS Rhodopseudomonas palustris.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodopseudomonas.
OX NCBI_TaxID=1076;
XP [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CGA009 / ATCC BAA-98;
RX PubMed=14704707;
RA Laimier F.W., Chain P., Hauser L., Lamerdin J., Malfatti S., Do L.,
RA Land M.L., Pelletier D.A., Beatty J.T., Lang A.S., Tabita F.R.,
RA Gibson J.L., Hanson T.E., Bobst C., Torres y Torres J.L., Peres C.,
RA Harrison F.H., Gibson J., Harwood C.S.;
RT "Complete genome sequence of the metabolically versatile
RT photosynthetic bacterium Rhodopseudomonas palustris.";
RL Nat. Biotechnol. 22:55-61(2004).
RR EMBL; BX572598; CAE27228.1;

```

RC STRAIN=ADP1;
RX MEDLINE=94237485; PubMed=8181753;
RA Hartnett G.B., Ornston L.N.;
RT "Acquisition of apparent DNA slippage structures during extensive
RT evolutionary divergence of *pcdA* and *catD* genes encoding identical
RT catalytic activities in *Acinetobacter calcoaceticus*.";
RL Gene 142:23-29(1994).
RN [4]
RP SEQUENCE FROM N.A.
RP STRAIN=ADP1;
RC MEDLINE=94341565; PubMed=8063101;
RX Kowalchuk G.A., Hartnett G.B., Benson A., Houghton J.E., Ngai K.L.,
RA Ornston L.N.;
RT "Contrasting patterns of evolutionary divergence within the
RT *Acinetobacter calcoaceticus* *pcA* operon";
RL Gene 146:23-30(1994).
RN [5]
RP SEQUENCE FROM N.A.
RP STRAIN=ADP1;
RX MEDLINE=95095936; PubMed=8002591;
RA Elasmore D.A., Ornston L.N.;
RT "The *pcA*-*poB* supraoperonic cluster of *Acinetobacter calcoaceticus*
RT contains *guilA*, the structural gene for *guinate*-shikimate
RT dehydrogenase.";
RL J. Bacteriol. 176:7659-7666(1994).
RN [6]
RP SEQUENCE FROM N.A.
RP STRAIN=ADP1;
RC MEDLINE=93194074; PubMed=8449410;
RX DiMarco A.A., Averhoff B.A., Kim E.E., Ornston L.N.;
RA "Evolutionary divergence of *pobA*, the structural gene encoding *p*-
RT hydroxybenzoate hydroxylase in an *Acinetobacter calcoaceticus* strain
RT well-suited for genetic analysis";
RL Gene 125:25-33(1993).
RN [7]
RP SEQUENCE FROM N.A.
RP STRAIN=ADP1;
RX MEDLINE=93322329; PubMed=8331077;
RA DiMarco A.A., Averhoff B., Ornston L.N.;
RT "Identification of the transcriptional activator *pobR* and
RT characterization of its role in the expression of *pobA*, the structural
RT gene for *p*-hydroxybenzoate hydroxylase in *Acinetobacter*
RT *calcoaceticus*.";
RL J. Bacteriol. 175:4499-4506(1993).
RN [8]
RP SEQUENCE FROM N.A.
RP STRAIN=ADP1;
RX MEDLINE=22401456; PubMed=12514037;
RA Smith M.A., Weaver V.B., Young D.W., Ornston L.N.;
RT "Genes for chlorogenate and hydroxycinnamate catabolism (*hca*) are
RT linked to functionally related genes in the *dca*-*pcA*-*guil*-*pob*-*hca*
RT chromosomal cluster of *Acinetobacter* sp. strain ADP1.";
RL Appl. Environ. Microbiol. 69:524-532(2003).
RN [9]
RP SEQUENCE FROM N.A.
RP STRAIN=ADP1;
RC Parke D., Ornston L.N.;
RT "Hydroxycinnamate (*hca*) catabolic genes from *Acinetobacter* sp. strain
RT ADP1 are repressed by *HcaR* and induced by hydroxycinnamoyl-CoA
RT thioesters.";
RL Appl. Environ. Microbiol. 0:0-0(2003).
RN DR EMBL; L05770; AA154850.1; -;
DR HSSP; P08659; ILCT.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 1.
RW ligase.
SQ SEQUENCE 626 AA; 70102 MW; 3C641E83ABD0CDA CRC64;

Query March 52.6%; Score 1621.5; DB 2; Length 626;
Match local similarity 53.0%; Pred. No. 1e-109;
Matches 308; Conservative 96; Mismatches 174; Indels 3; Gaps 3

Qy	5	EALLPFGRIILIERLEHNAKTRPEOCTVAAAPAAENEMPRISVAEMFHNVRALAOSSLPGI	64
Db	37	EQLAAVPEKLLDLRLIHFAKVPYHTFPAKRTODEWILSLTAESTLQRAWHIAQALHQRNL	96
Qy	65	SAERPLLIVSNDNDLEHLQALFAGANYAGIPYCPVSPAYSLSQDLAKLHIVGLQPGLVF	124
Db	97	SGEPRPIVVSNDLEHLTLTSMGALACGPFPAISPAVSLVSQDQGLKHVFDVLTIPGVVY	156
Qy	125	AADAAPOPRAIETTLPPDDVPAIFPRGELAGRRTVSPDSLLEQPGGIEADNAPATGPDTI	184
Db	157	ANDQAAKAIQSCSTODHIEIVTNKGIIIGDVCTSPFOSILDTP-VTVQEHYETLDEHQI	215
Qy	185	AKPLFTSGSTLPRAVPTQRMLCANQOMLQTFPVEEPRPVYVDMLPMHTFGSGSHNI	244
Db	216	AKPLFTSGSTLPRAVPTTHMLCLNQOMLQTFPEEDTPVLLDMLSMHTFGSHNV	275
Qy	245	GIVLVNGTIVYLDQKPTAQGAFTLRLNSEISPAVYLVPEKMBELVGALEPDSLTRE	304
Db	276	GIALYNGTIVYIDDKGPVPKGFDEITRLKESITPVLYNVPKGMBELTDALEKAELEDR	335
Qy	305	FFAMKLFPPFAAAGLSQGINDRLDRVAEQHGEKIRPMAGLGMTETAPSCITFTGPLSMA	364
Db	336	FFAVYKTLFFGAALSEAGNRRLDKIAQAHGEXIRIMSGIMETAPSCAFTTGPRVMA	395
Qy	365	GYIGLPAGCIVKLVYPVDKQLEGRFHGPHWGSYWR-APEONAOAFDEEGYSGDAITL	423
Db	396	GFIGYPAGCGIKLVPYPGDKLIEFCVRGKHMKGTWRLKADQOSTIFBDDGFFHGDVRL	455
Qy	424	ADPADPOKGLMFDGRIAEDPFLSGSPVSPGLSTRAVLEGSGYVLDVVAAPRECLGL	483
Db	456	VDPNPDPAQGLMYDGRIAEDPFLNTGTFVNVGTLTNKALIQNLIIQDVICISNLAIIGF	515
Qy	484	LVPRLLDCRLSLGIG-KEASDAVIVLASEPRAPRAMPDLKRLNEAQNASRIMVGLD	542
Db	516	LVPRLLDCAQDAFALNMKNHSAIDILQHPKVOQMPROGLIQYNQDATGSSNRVSMVLYMT	575
Qy	543	TPPSIDKGEVTDKGSINQRAVLQWRSAKVDAIYGEDQSMI	583
Db	576	EAPQDAGEVTDKGNLQNSNLTLRKRALVBEIYNKQTDNPL	616
RESULT 10			
Q6FBJ6 PRELIMINARY; PRT: 624 AA.			
Q6FBJ6			
AC	Q6FBJ6		
DT	05-JUL-2004 (TREMBLrel. 27, Created)		
DT	05-JUL-2004 (TREMBLrel. 27, Last sequence update)		
DT	05-JUL-2004 (TREMBLrel. 27, Last annotation update)		
DE	Coenzymase A ligase (EC 6.2.1.3).		
OS	Name=hcaac; Order=edlocuNames=CiCIND1724;		
OC	Acinetobacter sp. (strain ADP1).		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
OC	Moraxellaceae; Acinetobacter.		
OX	NCBI_TaxID=62977;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RA	Barbe V., Valleneet D., Fonknechten N., Kreilmeyer A., Ozias S.,		
RA	Labarre L., Cruveiller S., Robert C., Duprat S., Wincker P.,		
RA	Ornston L.N., Weissenbach J., Matilliere P., Cohen G.N., Medigue C.,		
RT	"Unique features revealed by the genome sequence of Acinetobacter sp.		
RT	ADP1, a versatile and naturally transformation competent bacterium.";		
RL	Nucleic Acids Res. 0:0-0(2004).		
DR	EMBL; CR543861; CAG68566.1; -		
DR	InterPro; IPR000873; AMP-bind.		
DR	Pfam; PF00501; AMP-binding; 1.		
KW	Complete proteome; Ligase.		
SO	SEQUENCE 624 AA; 69856 MW; 724419C9BA506580 CRC64;		

Query Match	52.5%	Score 1618.5;	DB 2;	Length 624;
Best Local Similarity	52.8%;	Pred. No. 1.7e-109;		
Matches 307;	Conservative 97;	Mismatches 174;	Indels 3;	Gaps 3;

QY 184 IAKFLFTSGSTKLPAKAVPTTQRMLCANQOMLQTFPVFGSEPPVLVDMLPMNHTFGSSHN 243
Db 217 IAKFLFTSGSTKLPAKAVTTQGMLCANQOMLQTFPEFAEVPVLVDMLPMNHTFGSSHN 276
QY 244 IGVLYNGGTYIYDDGKPTAQGFPAETLRNLSISPTLYLVTPKGMELVGLERDSTLRE 303
Db 277 VGVLYNGGTYIYDDGKPTAQGFPAETLRNLSISPTLYLVTPKGMELVGLERDSTLRE 336
QY 304 RPFARMKLPFPAAGSOGIMDRLDRAVEOHCGERIMMAGLGTETAPSCFTTGPLSM 363
Db 337 RPFARMKLPFPAAGSOGIMDRLDRAVEOHCGERIMMAGLGTETAPSCFTTGPLSLV 396
QY 364 AGYIGLPAPCEVYLVPDCKLEGRFPGHVMGSGYMPAPRONAQAPEDEGYCGDAIKL 423
Db 397 AGYIGLPAPCEVYLVPDCKLEGRFPGHVMGSGYMPAPRONAQAPEDEGYCGDAVRL 456
QY 424 ADPAPOKGLMPDRIEDKRLSGVVSVPGLTRAVLVBGGSYVLDDVVAAPRECLGL 483
Db 457 CDEGNPQGLMPDRIEDKRLSGVVSVPGLTRAVLVBGGSYVLDDVVAAPRECLGL 515
QY 484 LVPEPRLDICALSGLEKASDAEVLASEPRAWFADMLKRLNREATGNASRIMVVGLLDT 543
Db 516 LVPEPRLDICALSGLEKASDAEVLASEPRAWFADMLKRLNREATGNASRIMVVGLLDT 575
QY 544 PPSIDKGEVTDKGSINQRAVLQWRSAKVDALYRG 577
Db 576 APPIDRGEITDKGSINQRAVLQWRSAKVDALYRG 609

RESULT 8

Q8XT88 PRELIMINARY; PRT; 624 AA.

ID Q8XT88
AC Q8XT88
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PROBABLE FERULOYL-COA SYNTHETASE PROTEIN.
GN Name=fcsl; Synonyms=RS05196; OrderedlocusNames=RS0227;
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Manganot S.,
RA Atilat M., Billault A., Broctier P., Camus J.C., Cateolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiek T.,
RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).
CC - - SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
EMBL: AL646077; CAD17378.1; -.
DR HSP; P08659; ILC1.
DR GO; GO:0003624; F: catalytic activity; IEA.
DR GO; GO:0008152; P: metabolism; IEA.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 1.
DR PRINTS; PR00154; AMPBINDING.
KW Complete proteome; Plasmid.
SQ SEQUENCE 624 AA; 67134 MW; 75EE2038ACFP25A1 CRC64;

Query Match 63.0%; Score 1940.5; DB 2; Length 624;
Best Local Similarity 63.7%; Pred. No. 5.1e-133;
Matches 373; Conservative 69; Mismatches 143; Indels 1; Gaps 1;

QY 1 MSLEALPPGRILERLEHMAKTRPQTCVAARAANGEWRRISYAEFHNRAIAQSL 60

Db 35 LRSAEPGAYBERMTCLVNGAGQHPDRVLAARGADGQERITTYAQMTRARAVGALL 94
QY 61 PYGSAERPLIYSGNDLEHLOAFGAMVAGIYPCVPSVYSLSDLAIRHVGLOP 120
Db 95 ARGISPERPLIISGNDLQHLQALGAMVAGIYPCVPSVYSLSDLAIRHVGLOP 154
QY 121 GLVPAADAPFORAETILPDDVPAIFTRGELAGRTVSFSLLEDPGIEADNAPATG 180
Db 155 GLVYATGALFAGALQAVVPEPTEVAIDHGEVAGRAVTRLASIL-ATGPADVDAANARVG 213
QY 181 PDIITAKFLFTSGSTKLPAKAVPTTQRMLCANQOMLQTFPVFGSEPPVLVDMLPMNHTFGG 240
Db 214 PDIITAKFLFTSGSTKSPKATYTHRMLCNQOMLQTFPGFGAPVLLDMLPMNHTFGG 273
QY 241 SHNIGIYNGGTYIYDDGKPTAQGFPAETLRNLSISPTLYLVTPKGMELVGLERDST 300
Db 274 SHNIGIYNGGTYIYDDGKPTAQGFPAETLRNLSISPTLYLVTPKGMELVGLERDST 333
QY 301 LRERFARMKLPFPAAGSOGIMDRLDRAVEOHCGERIMMAGLGTETAPSCFTTGP 360
Db 334 LRERFARMKLPFPAAGSOGIMDRLDRAVEOHCGERIMMAGLGTETAPSCFTTGP 393
QY 361 LSWAGTYGLPAPCEVYLVPDCKLEGRFPGHVMGSGYMPAPRONAQAPEDEGYCGDA 420
Db 394 LSWAGTYGLPAPCEVYLVPDCKLEGRFPGHVMGSGYMPAPRONAQAPEDEGYCGDA 453
QY 421 IXLADPADPOKGLMPDRIEDKRLSGVVSVPGLTRAVLVBGGSYVLDDVVAAPRECL 480
Db 454 IXLADPADPOKGLMPDRIEDKRLSGVVSVPGLTRAVLVBGGSYVLDDVVAAPRECL 513
QY 481 LGLVPEPRLDICALSGLEKASDAEVLASEPRAWFADMLKRLNREATGNASRIMVVG 540
Db 514 LGLVPEPRLDICALSGLEKASDAEVLASEPRAWFADMLKRLNREATGNASRIMVVG 573
QY 541 LDTPEPIDKGEVTDKGSINQRAVLQWRSAKVDALYRGEDSMRDE 586
Db 574 LDTPEPIDKGEVTDKGSINQRAVLQWRSAKVDALYRGEDSMRDE 619

RESULT 9

Q8RM01 PRELIMINARY; PRT; 626 AA.

ID Q8RM01
AC Q8RM01
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Coenzyme A ligase.
GN Name=hcaC;
OS Acinetobacter sp. (strain ADP1).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Acinetobacter.
OX NCBI_TaxID=62977;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ADP1;
RX MEDLINE=98175676; PubMed=9515921;
RA Geriecher U., Segura A., Ornstom L.N.;
RT "Peau, a transcriptional activator of genes for protocatechuate
RL utilization in Acinetobacter";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ADP1;
RX MEDLINE=90130333; PubMed=2298704;
RA Hartnett C., Neidle E.L., Ngai K.L., Ornstom L.N.;
RT "DNA sequences of genes encoding Acinetobacter calcoaceticus
RT protocatechuate 3,4-dioxygenase: evidence indicating shuffling of
RT genes and of DNA sequences within genes during their evolutionary
RL divergence";
RN [3]
RP SEQUENCE FROM N.A.

```

QY 61 PYGLSAPERLLIVSGNDLEHLQALFAGMAYAGIPYCVSPAYSLLSODLAKLRIHVGLDLP 120
DB 61 GLGLSAPERLLIVSGNDLEHLQALFAGMAYAGIYCVSPAYSLLSODLAKLRIHVGLDLP 120
QY 121 GLVFAADAPFORAIEITLIPDDVPALFTRGELAGRRVTSFDSLLEOPGCI- EADNAFAAT 179
DB 121 GIVFVAVSDSPQFQAFAPAVLDSDGVI SVNGQVAVGRHISFDSL- QPGLAADAFAAT 179
QY 180 GEDDTAKELFTSGSTYLPRAVPTTQMLCANQOMLQTFPPVGESEPPVLVDMLPMNHTTG 239
DB 180 GEDDTAKELFTSGSTYLPRAVPTTQMLCANQOMLQTFPPVGESEPPVLVDMLPMNHTTG 239
QY 240 GSHNIGIVLYNGSTYLLDQKPTAOGFAETRLNLSISPTAYLTVPKGWEIVGALERDS 299
DB 240 GSHNIGIVLYNGSTYLLDQKPTAOGFAETRLNLSISPTAYLTVPKGWEIVGALERDS 299
QY 300 TLREFFAPMKLFFFAAGLSQGIWDLRVAEOHCGERI RMAGLGMTAASCFTTGG 359
DB 300 ALREVEFAIILKLFPAAGLSQSVMDRLRIAQHGGERIRMAAGLGMTAASCFTTGG 359
QY 360 PLSMAGYIGLPAPGCEVKLVVDGKLEGRFPHVMSGYRABEQNAQAFDEGGYICSGD 419
DB 360 PLSMAGYIGLPAPGCEVKLVVDGKLEGRFPHVMSGYRABEQNAQAFDEGGYICSGD 419
QY 420 ATKLADPADPOKGLMFDGRIADPFKLSGCVFVSGPLRTRAVLGGSYVLDDVVAAPDRE 479
DB 420 ATKLADPADPOKGLMFDGRIADPFKLSGCVFVSGPLRTRAVLGGSYVLDDVVAAPDRE 479
QY 480 CIGLVLFPRLDGRALSGLGEKESDAEVLASEPVRAMFADWLKRLNREATGNSRIMVYG 539
DB 480 CIGLVLFPRLDGRALSGLGEKESDAEVLASEPVRAMFADWLKRLNREATGNSRIMVYG 539
QY 480 CIGLVLFPRLDGRALSGLGEKESDAEVLASEPVRAMFADWLKRLNREATGNSRIMVYG 539
DB 480 CIGLVLFPRLDGRALSGLGEKESDAEVLASEPVRAMFADWLKRLNREATGNSRIMVYG 539
QY 540 LLDTPPSIDKGEVTDKGSINQRAVLQWRSAKVDALYRGDSML 583
DB 540 LLDTPPSIDKGEVTDKGSINQRAVLQWRSAKVDALYRGDSML 583
QY 540 LLDTPPSIDKGEVTDKGSINQRAVLQWRSAKVDALYRGDSML 583
DB 540 LLDTPPSIDKGEVTDKGSINQRAVLQWRSAKVDALYRGDSML 583

```

RESULT 6

```

QY 070V06 PRELIMINARY; PRT; 618 AA.
ID 070V06;
AC 070V06;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Feruloyl-CoA synthetase.
GN Name=ICB;
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=T2;
RA Ficca A.G., Di Gioia D., Barghini P., Fava F., Ruzzi M.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC -1- Similarity: Belongs to the ATP-dependent AMP-binding enzyme
family.
DR EMBL: AU536324; CAD60263.1; -.
DR InterPro: IPR000873; AMP-bind.
DR Pfam: PF00501; AMP-binding.1.
DR PRINTS: PRO0154; AMPBINDING.
SQ SEQUENCE 618 AA; 66992 MW; 57EBEB41166EBD33 CRC64;

```

```

Query Match 68.2%; Score 2103; DB 2; Length 618;
Best Local Similarity 71.4%; Pred. No. 6,8e-145;
Matches 410; Conservative 48; Mismatches 114; Indels 2; Gaps 2;

```

```

QY 4 LEALLPFGRIILERLHMAKTRPEQTCVABANGERRISYAEFMHNVAIAQSLLPYG 63
DB 38 VETLIDYVPALMERLLHMAQORSEQTFVARRADGQOHITVAAMLQVRVRIATWLLGOG 97
QY 64 LSAERPLLIVSGNDLEHLQALFAGMAYAGIPYCVSPAYSLLSODLAKLRIHVGLDPLV 123
DB 64 LSAERPLLIVSGNDLEHLQALFAGMAYAGIPYCVSPAYSLLSODLAKLRIHVGLDPLV 123

```

```

DB 98 LSAERPLLIVSGNDLEHLQALFAGMAYAGIPYCVSPAYSLLSODFGLKRIHVGLDPLV 157
QY 124 PAADAPFORAIEITLIPDDVPALFTRGELAGRRVTSFDSLLEOPGCI- EADNAFAATGPT 183
DB 158 PAADAA-YGRAIDAVEEPSINIVLGGELAGNASTPFAVLTVAQPCADQAFALRPDS 216
QY 184 IAKFLFTSGSTYLPKAVPTTQMLCANQOMLQTFPPVGESEPPVLVDMLPMNHTFGSHN 243
DB 217 LAKFLFTSGSTYLPKAVPTTQMLCANQOMLQTFPPVGESEPPVLVDMLPMNHTFGSHN 276
QY 244 IGIVLYNGSTYLLDQKPTAOGFAETRLNLSISPTAYLTVPKGWEIVGALERDSTRE 303
DB 277 VGIVLYNGSTYLLDQKPTAOGFAETRLNLSISPTAYLTVPKGWEIVGALERDSTRE 336
QY 304 RFPAPMKLFFFAAGLSQGIWDLRVAEOHCGERI RMAGLGMTAASCFTTGGPLM 363
DB 337 RFPAPMKLFFFAAGLSQGIWDLRVAEOHCGERI RMAGLGMTAASCFTTGGPLM 396
QY 364 AGYIGLPAPGCEVKLVVDGKLEGRFPHVMSGYRABEQNAQAFDEGGYICSDAIXL 423
DB 397 AGYIGLPAPGCEVKLVVDGKLEGRFPHVMSGYRABEQNAQAFDEGGYICSDAIXL 456
QY 424 ADPADPOKGLMFDGRIADPFKLSGCVFVSGPLRTRAVLGGSYVLDDVVAAPDRECLG 483
DB 457 CDRGNPQLGLMFDGRIADPFKLSGCVFVSGPLRTRAVLGGSYVLDDVVAAPDRECLG 515
QY 484 LVFRLDGRALSGLGEKESDAEVLASEPVRAMFADWLKRLNREATGNSRIMVGLDLP 543
DB 516 LVFRLDGRALSGLGEKESDAEVLASEPVRAMFADWLKRLNREATGNSRIMVGLDLP 575
QY 544 PSIDKGEVTDKGSINQRAVLQWRSAKVDALYRG 577
DB 576 APIDNGEITDKGSINQRAVLQWRSAKVDALYRG 609

```

RESULT 7

```

QY 070V06 PRELIMINARY; PRT; 618 AA.
ID CAD60263
AC CAD60263;
DT 02-MAR-2004 (T-EMBLrel. 27, Created)
DT 02-MAR-2004 (T-EMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (T-EMBLrel. 27, Last annotation update)
DE Feruloyl-CoA synthetase.
GN Name=ICB;
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=T2;
RA Ficca A.G., Di Gioia D., Barghini P., Fava F., Ruzzi M.;
RL "Identification of Pseudomonas putida strain T2 genes involved in
RT ferulate catabolism.";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AU536324; CAD60263.1; -.
SQ SEQUENCE 618 AA; 66992 MW; 57EBEB41166EBD33 CRC64;

```

```

Query Match 68.2%; Score 2103; DB 2; Length 618;
Best Local Similarity 71.4%; Pred. No. 6,8e-145;
Matches 410; Conservative 48; Mismatches 114; Indels 2; Gaps 2;

```

```

QY 4 LEALLPFGRIILERLHMAKTRPEQTCVABANGERRISYAEFMHNVAIAQSLLPYG 63
DB 38 VETLIDYVPALMERLLHMAQORSEQTFVARRADGQOHITVAAMLQVRVRIATWLLGOG 97
QY 64 LSAERPLLIVSGNDLEHLQALFAGMAYAGIPYCVSPAYSLLSODLAKLRIHVGLDPLV 123
DB 98 LSAERPLLIVSGNDLEHLQALFAGMAYAGIPYCVSPAYSLLSODLAKLRIHVGLDPLV 157
QY 124 PAADAPFORAIEITLIPDDVPALFTRGELAGRRVTSFDSLLEOPGCI- EADNAFAATGPT 183
DB 158 PAADAA-YGRAIDAVEEPSINIVLGGELAGNASTPFAVLTVAQPCADQAFALRPDS 216

```

Qy		181	PDLTAKLEFETSGSTGLPKRAVPTOTMLCANOOMLLOTPEVFSEEPVLYVDMLPNNHTFGG	240
Dd		181	PDSIAKLFETSGSTGLPKRAVITTOYMLCANOOMLLOTFEVPSEEPVLVDMLPMNHTFGG	240
Qy		241	SHNIGIVLYNSGTYYLLDGGKPTAOGFAETLRNLSEISPAVALYTPVKGMELVGALERDST	300
Dd		241	SHNGIVLYNSGTFFLDBSGKPTAOGFAETLRNLKSIPVALYTPVKGMELVNALREQDAE	300
Qy		301	LRERPFAAMKLFPFAAGLSGCIWDRLDRLVAEOHCGGERIRMMAGLGMTATBASCFTTTGP	360
Dd		301	LRECFEFKMSTSPFFFAAGLSGSVMRDLRLVAEQHCGGERIRMMAGLGMTAAASCTFTTGP	360
Qy		361	LSMAGYIGLPAPGCCVTKLVPDVDTKEGRPHGBHNWSGYRARAEQNQAEDEGCYCSSGA	420
Dd		361	LSMAGYIGLPAPGCCVTRLVPDVDTKEGRRGPHIMGVRAAQQTAEVDDAGPYCSGDA	420
Qy		421	IKLADPADPOKGLMPDGRIAEDEFKLSGCVFVSGPLRTAVLEGSSYVLDVVVAAPDREC	480
Dd		421	IKLADPADPOKGLMPDGRIAEDFKLSGCVFVSGLPRNRAYLEGAPYODLVVAAPDREC	480
Qy		481	LGILLVPPRLDCRALSGIAGEASDAEVLAESFVRAMPADMILKRLREATGNASRIIMVGL	540
Dd		481	LGAILVPPRLTYECRRLAGIAGEASEAQVLAASAVRWOFGDWLQCINREASGNASRIEMIAL	540
Qy		541	LDTPPSIDKGVTDKGSTNOARVYLQWRRAKVVALTRGEOSMLRDE	596
Dd		541	QVEPASIDRGEITTDKGSINORAVILOWRABOVENLYRGREPSTLRAB	596

RESULT 4
CAD60268 PRELIMINARY; PRT; 589 AA.

CAD60268
AC CAD60268;
DT 02-MAR-2004 (TREMBLrel. 27, Created)
DT 02-MAR-2004 (TREMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBLrel. 27, Last annotation update)
DE Feruloyl-CoA synthetase.
GN FCS.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=294;
RX [1]
RP SEQUENCE FROM N.A.
RC STAIN=BFI3;
RA Barghini P., Civolani C., Ficca A.G., Schlessner A., Ruzzi M.;
RT "Cloning and Characterization of Ferulate Catabolism Genes from
Pseudomonas fluorescens BFI3."
RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
RM EMBL; AJ563635; CAD60268.1; -.
SQ SEQUENCE 589 AA; 64311 MW; ECDCC7D40EDD55EB2B CRC64;

	Query Match	79.5%	Score 2449	DB 21	Length 589
	Best Local Similarity	77.8%	Pred. No. 3.4e-170		
	Matches 456	Conservative 54	Mismatches 76	Indels 0	Gaps 0
QY	1	MRSLEALLPFFGRILRLERLHMAKTRPEQTCVAARAANGEMRISYAEMFNHVAIAQSILL	60		
Db	1	MRSLEKPIALALDRLDLRLVHMAVRPEQTFIARAGCDMRLVSYGMDLSVRAIAQSILL	60		
QY	61	PYGSABRPILLIVSGNDLEHLQAFNGAYAGIPYCPVSPAYSLISODLAKLPHIYGLIOP	120		
Db	61	PYGSADPPLLLISGNDLEHLQALAGNAYAGIPYCPVSPAYSLISODPRAKLHVCDLIOP	120		
QY	121	GLVEAADAPFORAIEITLLPDDVPAIFMRGEIAGRRTSPSLIEQPGIEADNNAFAATG	180		
Db	121	GLVEVSDASAVQRAIDAVLPEPETPLISVRSQVPGRRQASPSLILAPGGAEDADNAFAATG	180		
QY	181	PDTIAKFLFTSGSTKLPAKAVPTTQAMLCANQOMLIQTTPVVGEEPPVLVDMLPMNHTGG	240		
Db	181	PDSIAKFLFTSGSTKLPAKAVITTTQRMLCANQOMLIQTTPVVGEEPPVLVDMLPMNHTGG	240		

QY	241	SHNIGVLYNAGTYXLLDQCKPTAOCFAETLRULSISIPAYITVPGMEVLGALRDSPT	300
Db	241	SHNIGVLYNAGTYXLLDQCKPTAOCFAETLRULSISIPAYITVPGMEVLGALRDSPT	300
QY	301	LREFFFAFMKLFPPAAGLSOGIMWRDLRVAEONHGERIRMMAGLGMTETASCTFTTGP	360
Db	301	LRECFEFKMSLFPPAAGLSQSVMDRLRVAEQHGERIRMMAGLGMTETASCTFTTGP	360
QY	361	LSMAGYIGLPAPEGCEVTKLVPVDGKLEGRFHGPHVNSGYWRADEQUNQAIFDEEGYICSGDA	420
Db	361	LSMAGYIGLPAPEGCEVRLVPVDGKLEGRFPGPHIMPYWRAAQQAFAEVDAQFCYCSGDA	420
QY	421	IKLADPADPQKGLMFDPGRIAEDFKLSSGVFVSVGLRTRAVYLEGGSYVLVDVVAAPDRRC	480
Db	421	IKLADPADPQKGLMFDPGRIAEDFKLSSGVFVSVGLRNRVAILEGAPYVDLVLVAAPDRRC	480
QY	481	IGLLVFPRLDCCRLSLGIGKESASDAEVLASBEVRAMPFMDMLKRLREATGNSASRIWVGJ	540
Db	481	IGALLVFPRLYECRRLAGIGAESAEQVILASAVRQWFGMWLQCLNREASGNSASRLIEWIAL	540
QY	541	LDPPTIDKGEVTDKGSINORAVLQMRSAKVDAIYRGEDQSMTLDE	586
Db	541	QVEPASIDRGEITIDKGSINQRAVLQMRBEQVENLIRGREPSLIRPAE	586

RESULT 5	098HK0	PRELIMINARY;	PRT;	589 AA.
ID	Q98HK0			
AC	Q98HK0;			
DT	01-JUN-2003 (TREMBLrel. 24, Created)			
DT	01-JUN-2003 (TREMBLrel. 24, Last sequence update)			
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)			
DE	PeruIoyl-COA-synthetase.			
GN	Name=Ics; OrderedlocusNames=PP3356;			
OS	Pseudomonas putida (strain KT2440).			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;			
OC	Pseudomonadaceae; Pseudomonas.			
OX	NCBI_TaxId=160488;			
UN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=22432060; PubMed=12534463;			
RA	Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Halbert H.,			
RA	Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,			
RA	Brinkac L.M., Bauman M.J., Deboy R.T., Daugherty S.C., Kolonay J.F.,			
RA	Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,			
RA	Hance I., Chris Lee P., Holtzapfe E.K., Scanlan D., Tran K.,			
RA	Mozzaz A., Uitterlind T.R., Rizzo M., Lee K., Kosack D., Moestl D.,			
RA	Weidler H., Lauber J., Stjepandic D., Hohnel J., Straetz M., Heim S.,			
RA	Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoef A., Tsemmler B.,			
RA	Fraser C.M.;			
RT	"Complete genome sequence and comparative analysis of the			
RT	metabolically versatile Pseudomonas putida KT2440.";			
RL	Environ. Microbiol. 4:799-808(2002).			
CC	-1-SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme			
CC	family.			
CC				
DR	EMBL; AF016786; AAM68960.1; -			
DR	TIGR; PP3356; -			
DR	GO; GO:0003824; P:catalytic activity; IEA.			
DR	GO; GO:0008152; P:metabolism; IEA.			
DR	InterPro; IPR000873; AMP-Bind.			
DR	Pfam; PF00501; AMP-binding; 1.			
DR	PRINTS; PR00154; AMPBINDING.			
DR	Complete proteome.			
KW	SEQUENCE 589 AA; 64283 MW; 4B02BC8F47253237 CRC64;			

	Query Match	76.2%;	Score 2349;	DB 2;	Length 589;
	Best Local Similarity	75.9%;	Pred. No. 6-9e-163;		
	Matches 443;	Conservative	56;	Mismatches 83;	Indels 2; Gaps 2
Oy	1 MRSLEALLPPFGRIILERLHEHWAKTRPEECTCYAAPAAANGEMRISIAEHFHVAVRAIQSLI	60			
	: :				
ob	1 MQDPESLPAPCPARLIERLVHWAQVRPTDTEFAAQADCAWMSISVQNLAVRTIAANLI	60			

```

QY 361 LSMAGYIGLPAPGCEVKLVVDGKLEGRFGHVMGYPAPRONQAPEEGYCSGA 420
DB 361 LSMAGYIGLPAPGCEVKLVVDGKLEGRFGHVMGYPAPRONQAPEEGYCSGA 420
QY 421 IKLADPADPOKGLMFGDRIAEDEPKLSSGVFVSGLPTRAVLFGSGVLDVVAAAPDREC 480
DB 421 IKLADPADPOKGLMFGDRIAEDEPKLSSGVFVSGLPTRAVLFGSGVLDVVAAAPDREC 480
QY 481 LGLVFPRLDCCRALSGLGKEASDAEVLASEPYRAWFADMLKELNREATGNASRIWVGL 540
DB 481 LGLVFPRLDCCRALSGLGKEASDAEVLASEPYRAWFADMLKELNREATGNASRIWVGL 540
QY 541 LDTPEIDGSEVTDKGSINQRAVLQWRSAKVDAALYGEQDSMLRDEATL 589
DB 541 LDTPEIDGSEVTDKGSINQRAVLQWRSAKVDAALYGEQDSMLRDEATL 589

RESULT 2
Q881F2 PRELIMINARY; PRT; 589 AA.
ID 0881F2
AC 0881F2, 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Feruloyl-CoA synthetase.
GN Name=fcg, OrderedLocuNames=PSPTO2941;
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
RA Buehl C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA Gwin M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F.,
RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,
RA Nelson W.C., Davidsen T.M., Tran B., Russell D., Berry K.J.,
RA Khoult H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
RA Utterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,
RA Delaney T.P., Lazrowitz S.G., Martin G.B., Schneider D.J., Tang X.,
RA Bender C.L., White O., Fraser C.M., Collier A.;
RT "The complete genome sequence of the Arabidopsis and tomato pathogen
RT Pseudomonas syringae pv. tomato DC3000."
RT Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
DR EMBL: AE016866; AAC056434.1; -.
DR HSSP: P08659; 1LCT.
DR TIGR: PSP02941; -.
DR GO: GO:0003824; F: catalytic activity; IEA.
DR GO: GO:0008152; P: metabolism; IEA.
DR InterPro: IPR000873; AMP-bind.
DR Pfam: PF00501; AMP-binding; 1.
DR PRINTS: PR00154; AMPBINDING.
DR Complete proteome.
SQ SEQUENCE 589 AA; 64272 MW; FA135BCB9722AF12 CRC64;

Query Match 80.3%; Score 2474; DB 2; Length 589;
Best Local Similarity 78.9%; Pred. No. 5, 1e-172;
Matches 460; Conservative 50; Mismatches 73; Indels 0; Gaps 0;
QY 1 MSLLEALLPFGGILRLERLHMAKTRPEQTCVAAARAANGERRISYAEFHNVAIAQSLL 60
DB 1 MSLLEALLPFGGILRLERLHMAKTRPEQTCVAAARAANGERRISYAEFHNVAIAQSLL 60
QY 61 PYGLSARPLLIVSGNDLHLQLAFGANVAGIPYCVSPAYSLISODLAKLRHIVGLTOP 120
DB 61 PYGLSARPLLIVSGNDLHLQLAFGANVAGIPYCVSPAYSLISODLAKLRHIVGLTOP 120
QY 61 AYGLSARPLLIVSGNDLHLQLAFGANVAGIPYCVSPAYSLISODLAKLRHIVGLTOP 120
DB 61 AYGLSARPLLIVSGNDLHLQLAFGANVAGIPYCVSPAYSLISODLAKLRHIVGLTOP 120
QY 121 GLVFADAAAPFORAIEITILPDVPAIFTRGELAGRTVSFDSILLEPGGIEADNAFAATG 180
DB 121 GLVFADAAAPFORAIEITILPDVPAIFTRGELAGRTVSFDSILLEPGGIEADNAFAATG 180

```

```

DB 121 GLVFADAAAPFORAIEITILPDVPAIFTRGELAGRTVSFDSILLEPGGIEADNAFAATG 180
QY 181 PDTIAKFLFTSGSTKLPKAVPTTORMLCANQOMLOTFFVGESEBPVLYDMLPMNHTFGG 240
DB 181 PDTIAKFLFTSGSTKLPKAVPTTORMLCANQOMLOTFFVGESEBPVLYDMLPMNHTFGG 240
QY 241 SHNIGIVLVNGGIVYLDGKPTAGQFAETLRNLSISPAVYLVPPGMEELVGLRRDST 300
DB 241 SHNIGIVLVNGGIVYLDGKPTAGQFAETLRNLSISPAVYLVPPGMEELVGLRRDST 300
QY 301 LRRFPFAKMLPEFPAAGLSOGIWDRLDVAEQHCGERIRMAAGLQMTETAPSCITTTGP 360
DB 301 LRRFPFAKMLPEFPAAGLSOGIWDRLDVAEQHCGERIRMAAGLQMTETAPSCITTTGP 360
QY 361 LSMAGYIGLPAPGCEVKLVVDGKLEGRFGHVMGYPAPRONQAPEEGYCSGA 420
DB 361 LSMAGYIGLPAPGCEVKLVVDGKLEGRFGHVMGYPAPRONQAPEEGYCSGA 420
QY 421 IKLADPADPOKGLMFGDRIAEDEPKLSSGVFVSGLPTRAVLFGSGVLDVVAAAPDREC 480
DB 421 IKLADPADPOKGLMFGDRIAEDEPKLSSGVFVSGLPTRAVLFGSGVLDVVAAAPDREC 480
QY 481 LGLVFPRLDCCRALSGLGKEASDAEVLASEPYRAWFADMLKELNREATGNASRIWVGL 540
DB 481 LGLVFPRLDCCRALSGLGKEASDAEVLASEPYRAWFADMLKELNREATGNASRIWVGL 540
QY 541 LDTPEIDGSEVTDKGSINQRAVLQWRSAKVDAALYGEQDSML 583
DB 541 LDTPEIDGSEVTDKGSINQRAVLQWRSAKVDAALYGEQDSML 583

RESULT 3
Q70V01 PRELIMINARY; PRT; 589 AA.
ID Q70V01
AC Q70V01, 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Feruloyl-CoA synthetase.
GN Name=fcg;
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BF13;
RA Barghini P., Civolani C., Ficca A.G., Schiesser A., Ruzzi M.;
RT Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
DR EMBL: AU536325; CAD60268.1; -.
DR InterPro: IPR000873; AMP-bind.
DR Pfam: PF00501; AMP-binding; 1.
DR PRINTS: PR00154; AMPBINDING.
SQ SEQUENCE 589 AA; 64311 MW; ECDCTD40BDD55F2B CRC64;

Query Match 79.5%; Score 2449; DB 2; Length 589;
Best Local Similarity 77.8%; Pred. No. 3, 4e-170;
Matches 456; Conservative 54; Mismatches 76; Indels 0; Gaps 0;
QY 1 MSLLEALLPFGGILRLERLHMAKTRPEQTCVAAARAANGERRISYAEFHNVAIAQSLL 60
DB 1 MSLLEALLPFGGILRLERLHMAKTRPEQTCVAAARAANGERRISYAEFHNVAIAQSLL 60
QY 61 PYGLSARPLLIVSGNDLHLQLAFGANVAGIPYCVSPAYSLISODLAKLRHIVGLTOP 120
DB 61 PYGLSARPLLIVSGNDLHLQLAFGANVAGIPYCVSPAYSLISODLAKLRHIVGLTOP 120
QY 61 RYGLSARPLLIVSGNDLHLQLAFGANVAGIPYCVSPAYSLISODLAKLRHIVGLTOP 120
DB 61 RYGLSARPLLIVSGNDLHLQLAFGANVAGIPYCVSPAYSLISODLAKLRHIVGLTOP 120
QY 121 GLVFADAAAPFORAIEITILPDVPAIFTRGELAGRTVSFDSILLEPGGIEADNAFAATG 180
DB 121 GLVFADAAAPFORAIEITILPDVPAIFTRGELAGRTVSFDSILLEPGGIEADNAFAATG 180

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2005, 21:38:30 ; Search time 138 Seconds
(without alignments)
2455.766 Million cell updates/sec

Title: US-09-750-986D-30
Perfect score: 3082
Sequence: 1 MSLSLALPPGRIERLERH.....KVDALYRGEDQSMRLDEATL 589

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues
Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3082	100.0	589	2	Q9RLD9 pseudomonas
2	2474	80.3	589	2	Q881F2 pseudomonas
3	2449	79.5	589	2	Q70V01 pseudomonas
4	2449	79.5	589	2	CAD60268 pseudomonas
5	2349	76.2	589	2	Q88HK0 pseudomonas
6	2103	68.2	618	2	Q70V06 pseudomonas
7	2103	68.2	618	2	CAD60263 pseudomonas
8	1940.5	63.0	624	2	Q8XT88 ralsstonia s
9	1621.5	52.6	626	2	Q8RM01 acinetobact
10	1618.5	52.5	624	2	Q6FBJ6 acinetobact
11	1457	47.3	632	2	Q6N8W6 pseudomonas
12	1457	47.3	632	2	CAE27228 rhodospheu
13	1401	45.5	627	2	Q89CF2 bradyrhizob
14	1294	42.0	591	2	Q7CZAS agrobacteri
15	1294	42.0	611	2	Q8UF11 agrobacteri
16	1210.5	39.3	644	2	Q89VJ9 bradyrhizob
17	1129	36.6	623	2	Q89CJ5 bradyrhizob
18	1128	36.6	624	2	Q6N8Y3 rhodopseu
19	1128	36.6	624	2	CAE27148 rhodopseu
20	1002.5	32.5	596	2	Q9AB43 caulobacter
21	972	31.5	621	2	Q9A368 caulobacter
22	972	31.5	621	2	Q93KX9 pseudomonas
23	937	24.1	199	2	Q7NQT6 streptomyce
24	933	10.9	587	2	Q82B22 streptomyce
25	933	10.8	636	2	Q82B22 streptomyce
26	319.5	10.4	647	2	Q8KD98 chlorobiom
27	312.5	10.1	607	2	Q8G7J9 bifidobacte
28	298	9.7	476	2	Q50017 mycobacteri
29	290.5	9.4	642	2	Q9L0U1 streptomyce
30	280	9.4	600	2	Q7N131 photorhabdu
31	289	9.4	601	2	Q821G5 yersinia pe

32	289	9.4	601	2	AAS63793 yersinia
33	289	9.4	619	2	Q89UX1 bradyrhizob
34	288	9.3	658	2	Q8DJC6 synechococc
35	286	9.3	607	1	LCFH_HAMIN
36	284	9.2	700	2	Q8LKS5 arabidopsis
37	283.5	9.2	601	2	Q8AIV8 bacteroides
38	283.5	9.2	601	2	Q9KRP9 vibrio chol
39	281	9.1	612	2	Q82UB1 nitrosomona
40	281	9.1	701	2	Q9CSU7 arabidopsis
41	280.5	9.1	602	2	Q87SS1 vibrio para
42	280	9.1	604	2	Q6LV19 photobacter
43	280	9.1	604	2	CAG18856 photobact
44	279.5	9.1	473	2	Q73WT0 mycobacteri
45	279.5	9.1	473	2	AAS04897 mycobacte

ALIGNMENTS

RESULT 1
ID Q9RLD9 PRELIMINARY; PRT; 589 AA.
AC Q9RLD9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Feruloyl-CoA synthetase.
GN Name=fcgs;
OS Pseudomonas sp.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HR199;
RX MEDLINE=20011220; PubMed=10543794;
RA Overhage J., Frieleert H., Steinduechel A.,
RT "Biochemical and genetic analyses of the ferulic acid catabolism in
RT Pseudomonas sp. strain HR199."
RL Appl. Environ. Microbiol. 65:4837-4847(1999).
DR EMBL: AJ238746; CAB60226.1; -;
DR GO: GO:0003824; F: catalytic activity; IEA.
DR GO: GO:0008152; P: metabolic; IEA.
DR InterPro: IPR000873; AMP-bind.
DR Pfam: PF0501; AMP-binding; 1.
SQ SEQUENCE 589 AA; 64512 MW; 60C9A9B6AF787BCD CRC64;

Query Match 100.0%; Score 3082; DB 2; Length 589;
Best Local Similarity 100.0%; Pred. No. 2e-216;
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSLALPPGRIERLERHMAKTRPEOTCVARAANGERRISYAEPMHVAIAOSLT	60
Db	1	MSLALPPGRIERLERHMAKTRPEOTCVARAANGERRISYAEPMHVAIAOSLT	60
Qy	61	PVGLSARPLLIVSGNDLHQLAFGAMYAGIPYCPVSPAYSLISODLAKRHIVGLLOP	120
Db	61	PVGLSARPLLIVSGNDLHQLAFGAMYAGIPYCPVSPAYSLISODLAKRHIVGLLOP	120
Qy	121	GLVFADAAAPQPAITIIIPDVPAIFTRGELAGRTVSFDSLLBOPGIEADNAPATG	180
Db	121	GLVFADAAAPQPAITIIIPDVPAIFTRGELAGRTVSFDSLLBOPGIEADNAPATG	180
Qy	181	PDTIAKFLFTSGSTKLPKAVPTTQRLCANQQLQTFPVFGEPVLVDWLPWNHTFGG	240
Db	181	PDTIAKFLFTSGSTKLPKAVPTTQRLCANQQLQTFPVFGEPVLVDWLPWNHTFGG	240
Qy	241	SHNIGIVLYNGGYVLDDEKPTAOGFAETLRNLSEISPTAYLTVPGKMEIVGALBRDST	300
Db	241	SHNIGIVLYNGGYVLDDEKPTAOGFAETLRNLSEISPTAYLTVPGKMEIVGALBRDST	300
Qy	301	LRRFPFAKMLFFFAAGLSOGIWDRLDVAHQHCGERIRMAAGLMTETAPSCFTTGP	360
Db	301	LRRFPFAKMLFFFAAGLSOGIWDRLDVAHQHCGERIRMAAGLMTETAPSCFTTGP	360

This Page Blank (uspto)


```

Db      499 GCTGATACCTCTGTCGATGCTTGTGGGATCAAGTATGGTATGTTT----- 549
QY      93 ProTyrCysProValSerProAlaTyrSerLeu----- 104
Db      550 -----ACGTACTCTTACTTACTTCTTANCCAGATTAATGCGCAC 568
QY      105 -----SerGlnAspLeuAlaIysLeuArgHisIleValGlyLeuLeuGlnProGlyLeu 122
Db      589 AAGAGACGTCATGTTGTTAGAGGATCCAGATCGTCTGGAGAGATTTATGCAATATA 648
QY      123 ValPheAlaAlaAspAlaPro-----PheGln-ArgAlaIle 135
Db      649 CTGTCAATTCGAAAGTGAAGCCCTTGTGTGATTAACCTGAGTTTTCATCGCATTCG 708
QY      135 eGlnThrIleLeuProAspAspAlaProAlaIlePheThrArg-----GlyGlu----- 151
Db      709 GGAATCATTTTCTTACAGCGACAGCTCAAAATTTGTGACTCTTCTGGGGGGAATAATC 768
QY      152 -----LeuAlaGlyArgThrValSerPheAspSerLeuLeuGlnProGly 168
Db      769 GTCGTTGCTTACAGCGGGTAGGACACACAGCTCTATGTTACAGAAATTAAGAATT 828
QY      168 yGlyIleGlyAlaAspAsnAlaPheAlaIleThrGly----- 180
Db      829 TCGACAGAGAAACGTGCAAAATTTGCAAGATCTATGATTCGGAGATGATGATGTA 888
QY      181 -----ProAspThrIleAlaIysPheLeuPheThrSerGlySerThrIysLeuP 197
Db      889 ATATCATCATTCCTTACAGATATAGCCACAAATTATGATATGATCCAGATGAGACCAAGAAATCC 948
QY      197 OlyAlaValProThrThrGlnArgMetLeuCysAlaAsnGlnMetLeuLeuGln 217
Db      949 AAAAGGTATTTATCTACACATCAGAAATTTGTACACAGATTAAGAACTTGTCCGATTT 1008
QY      217 rPheProValPheGlyGluGluProValLeuValAspTyrPheProTyrPheHis 237
Db      1009 TGTGCTCT-----GCCGAAGCTGGGGAAGATTTCTGATATGTTGCCATCATGCGATGC 1062
QY      237 rPheGlyGlySer-----HisAsnIleGlyIle----- 246
Db      1063 TTATGAACGGGCTGTGAATACTCTCATTTTACATGTGAGTTCGACAAAGTATACGTC 1122
QY      246 ----- 246
Db      1123 TATAGATTTCTTAAGGATGATCTCAAGGTTATCAACGACACTATTTCACTTCC 1182
QY      247 -----ValLeuTyrAsnGlyGlyThrTyrTyrLeuAspAspGlyLysProTh 262
Db      1183 TTTAGTATATGAGACACTCTACAGCGGATTCAAAAGCAATTTCTGCAAGCTCCCTCG 1242
QY      262 rAlaGlnGlyPheAlaGluThrLeuArgMetLeuSerGluIleSerProThrAlaTyrIle 282
Db      1243 TCGTAAATTTTGGCCTTACATGATGATCAAAAGTCAC-----CTGCAATATAC 1290
QY      282 uThrValProLysGlyTyrGluGluLeuValGlyAlaLeuGluArgPheSer----- 299
Db      1291 GGAATGAAAGAGTTATGAGGCTCT-----TGTTTGACAAATAATCAAAAGCTCC 1344
QY      300 -----ThrLeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAl 315
Db      1345 AATGTATATGTTTCTGTTGGTATGTTGTTGGCGAGATGATTGTCATTTTTC----- 1399
QY      315 aAlaAlaGlyLeuSerGlnGlyIleThrAspArgLeuAspArgValAlaGluGlnHisCys 335
Db      1400 -----CTATGGCCA-----TTGCATATGTTGGCTGAAAGCTTGT 1434
QY      335 eGlyGluArgIleArgMetMetAlaGlyLeu----- 345
Db      1435 ACAAGAAATTTCTGTTGCTATTTGGATTAACAAGGCTGCTTAAAGTGAAGTGTAG 1494
QY      345 ----- 345

```

```

Db      1495 TTACCTATCATGTTTGACAAAGTTTTTTGAAGCCATCGGTGATATGTAACAAATGGGTA 1554
QY      346 -GlyMetThrGluThrAlaProSerCysThrPheThrGlnProLeuSerMetAlaGly 365
Db      1555 TGGTTTGACGAAACCTTACACGGTGTCTCTGGCGGAGGCTTAAGGTATACGTTCTGG 1614
QY      365 yTyrIleGlyLeuProAlaProGlyCysGluValLysLeuVal----- 379
Db      1615 CTCAGTTGGGCATCTTATTAAGATATACAGAGTTCAAAATTTGTATCATGATGAGCTGTAC 1674
QY      380 -----ProValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSerGly 397
Db      1675 TGTCTTCCACCTGTTCAAGGGCATGTCAAAAGTCAGAGGCCACCGGTATGAAAGG 1734
QY      397 yTyrTrpArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGlyTyrTyrCysSe 417
Db      1735 TTRCTCAAGAAATCCCTGCGCCACCAAGAGGTTATGACGATGATGATGATTCATAC 1794
QY      417 rGlyAspAlaIleLysLeuAlaAspProAlaAspProGlnLysGly----- 432
Db      1795 TGGAGATATGGGTGGATTACT-----CCTCAGCACTCAACAGAGCGAGTCG 1842
QY      433 -----LeuMetPheAspGlyArgIleAlaGluAspPheLysLeuSerSe 447
Db      1843 TAGCTGTGAGGTGTCATGCTTCTTGAAGCGCGTGCAGGAAGACT-ATCGTCTTTCAC 1901
QY      447 rGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGluGlyLyserty 467
Db      1902 AGGT-----GAAATGTGGAACATTGGAGATGTAAGAAGCGGCATGGAAGCAATTT 1955
QY      467 rValLeuAspValValAlaAlaAlaProAspArgLysCysLeuGlyLeuLeuValPheP 487
Db      1956 GATTCACAAATATAGTGTATTGAGACAGATCAACGCCCTTGAGACTATGTATCC 2015
QY      487 OArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGlyAla----- 502
Db      2016 A-----AACCAAGAAAGCAGACAGAGACGACGAAAGCAAAATTTACCTGTGTA 2066
QY      503 -SerAspAlaGluValLeuAlaSerGluProValArgAlaTyrPheAlaAspTyrLeu 522
Db      2067 TTCTGAATGCAATGAACCTTAGCAAGAGACGATTAACAGCATGCTTATGAAGAACTAG 2126
QY      522 sArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeuLeuAs 542
Db      2127 GAAATGAGCGTCACATGCTCGTTCAGATGGACCAAGTTCTG-----ATCGTGA 2177
QY      542 pThrProProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGlnArgAl 562
Db      2178 TGAACCATTCACGATAGACACGCTTATATGACACCGCAATGAATAAAGACGGAGCAA 2237
QY      562 aValLeuGlnTyrArgSerAlaLysValAspAlaLeuTyrArg 576
Db      2238 GGTGTGATCAATACAGAAATGAATAAGAAAGCTCTACAG 2280

```

Search completed: February 8, 2005, 02:57:24
 Job time : 4080 secs

```

QY 324 ---TrpAspArgLeuAspArgValAlaGlu---GlnHis-----CysGlyGluArgIle 339
DB 1345 TTCTATGAGCGCGCCCATGATGAGACACACACTTCTTCTGTGCTTCAACATC 1404
QY 340 ArgMetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGly 359
DB 1405 CGCTTGATGCGGGCTACGGCCCTCACTGACCTCA---GCCCCCACTTCATGTCAGT 1461
QY 360 ProLeuSerMetAlaGlyTyr---IleGlyLeuProAlaProGlyCysGluValIleLeu 378
DB 1462 CCTCAACTACCGCGCTGTACAGCTCAGCCAAAGTTGGTCCCGCGCTGCGGGAAGCTG 1521
QY 379 ValProValAspGlyIleLeuGluGlyArg-----PheHisGlyProHisValMetSer 396
DB 1522 GTAAACACAGACGACGAGGACATTGGCGAGANTGCTCTGTGGGCGCACCATATTTCATG 1581
QY 397 GlyTyrTrpArgAlaProGluGluAlaAlaGlnAlaPheAspGluGluGlyTyrCys 416
DB 1582 GGTACCTGAACATGAGAGCAAGACTTGTGAGCCATGACGAGAGAGGCTGCGCAC 1641
QY 417 SerGlyAspAlaIleIleLeuAlaAspProAlaAspProGluIleGlyLeuMetPheAsp 436
DB 1642 AGCGGATGATGCGCGCGCTG-----GACGCGATGCGCTTCTCATCATCACT 1689
QY 437 GlyArgIleAlaGluAspPheLeuSerSerGlyValPheValSerValGlyProLeu 456
DB 1690 GGGCGCTCAAGAAATTAATCATCACAGCTGTGGGAGAGANTGCCCCCTGTGCCATC 1749
QY 457 ArgThrArgAlaValLeuGluGlySerTyrValLeuAspValValAlaAlaAlaPro 476
DB 1750 GAGGAGCGCGTGAAGATGAG---CTGCCCATCATCAGCAAGCCATGCTCATTTGGGAC 1806
QY 477 AspArgGluCysLeuGlyLeuLeuValPhe-----ProArgLeuLeu 490
DB 1807 CAGAGAAAGTCTCTGCTCATGCTGCTCACCTTGAAGTGCATCTGAGACCACTCT 1866
QY 491 Asp-----CysArgAlaLeuSerGlyLeu 498
DB 1867 GACCAGACTGATTAATCTGACTGAACAAGCTGTGAGCTTGTCCAGAGGGGACAGCA 1926
QY 499 GlyLeuGluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAla 518
DB 1927 GCCACACACAGTGTCCAGATCATAGAGAAAGAGAGAGCCCTGTACACGCGCATCGAA 1986
QY 519 AspTrpLeuLeuArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpVal 538
DB 1987 GAGGGGATCCGAGGGGTCAACATGAACGCGCGCGCCCTACACACATCCAGAACTGG 2046
QY 539 GlyLeuLeuAspThrProProSerIleAspIleGlyValThrAspIleSerIle 558
DB 2047 GCCATTCTGAGAGACATCTTCTCATTTCCGGTGGAGAGTGGGTCCACGATGAACACTG 2106
QY 559 AsnGlnArgAlaValLeuGlnTrpArgSerAlaValAspAlaLeuTyrArgGlyGlu 578
DB 2107 AAACGGCTACAGCTTTTGGAGAGATCAAAAGATCATGACTCTCTTAC-----CAA 2160
QY 579 AspGlnSerMet 582
DB 2161 GAGCAAAAATG 2172

```

RESULT 15
CNSOA7CK 2366 bp mRNA linear HTC 06-FBB-2004
LOCUS Arabidopsis thaliana Full-length cDNA Complete sequence from clone
DEFINITION GSP1PGH76ZG05 of Hormone Treated Callus of strain col-0 of
Arabidopsis thaliana (thale cress).
ACCESSION BX824883.1 GI:42466280
VERSION 1
KEYWORDS HTC; GSP1; cDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

1 (bases 1 to 2366)
Castelli,V., Aubry,J.M., Jallion,O., Wincker,P., Clepet,C., Menard,M., Cruaud,C., Quetier,F., Scapellato,C., Schachter,V., Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2366)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted (18-NOV-2003) Genoscope (E-mail : seqref@genoscope.cns.fr
JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT
The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli V., Aubry J.M., Jallion O., Wincker P., Menard M., Cruaud C., Schachter V., Weissenbach J., Salanoubat M.
URGV INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich Information center for Protein Sequences). 5 prime and 3 prime are assembled with Phrap.
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full Length
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.

FEATURES
source
1..2366
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Col-0"
/db_xref="taxon:3702"
/clone="GSP1PGH76ZG05"
/issue_type="Hormone Treated Callus"
/plasmid="pCMVSPORT_6"
1..2366
/gene="At3g23790"

ORIGIN
Alignment Scores:
Pred. No.: 7,36e-12 Length: 2366
Score: 226.00 Matches: 151
Percent Similarity: 35.24% Conservative: 108
Best Local Similarity: 20.54% Mismatches: 263
Query Match: 7.33% Indels: 214
Gaps: 29
DB: 3

US-09-750-986D-30 (1-589) x CNSOA7CK (1-2366)

QY 2 ArgSerLeuGluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuIleTrp 21
DB 238 AAGAGATTAAAGCGCGGTTCGCATTC-----TTAGAAAGCTTA----- 276
QY 22 AlaLeuThrArgProGluGlnThrCysValAlaAlaAlaAlaAlaAlaAlaAlaAla 41
DB 277 -----TCATTGCCAAGGAG-----GCTGCTTGAGCTCTAATGATGAGAG 318
QY 42 -----ArgIleSerTyrAla 46
DB 319 TCTGCTCTGATATTGTGAGATCATCTGTGAGAACTACGATGACAGATGCGGTCGTA 378
QY 47 GluMetPheHisAsn-----ValArgAlaIleAlaGlnSerLeuLeu 60
DB 379 GATCCGATATCATGACCGCGCTTCTACATTCACGTACAGACAGTGGAAAGAAATCTTG 438
QY 61 ProTyr-----GlyLeuSerAlaGluArgProLeuLeuIleVal 73
DB 439 GACTTTGTTGAGGGTTTACAGACTGTTGAGATTAAAGACAGAGAGATTCGACTTTT 498
QY 74 SerGlyAsnAspLeuGluIleLeuGluIleAlaPheGly---AlaMetTyrAlaGlyIle 92

```

QY      338 gileArgMetMetAlaGlyLeu----- 345
Db      1659 TCtAAACATGGGCGGCTGCGGAGTGCGACGCTGATTCTCTCAGCGTAAACAT 1718
QY      346 -----GlyMetThrGluThrAlaProSerCysThrPheThrThrG1 359
Db      1719 CCTATCATGAGATTGATGACGCTTGGAG-----TGACGGGACTACACAC 1766
QY      359 yProLeuSerMetAlaGlyTyr-----IleGlyLeuProAlaProG1 373
Db      1767 AGTGTCACACCTGCGAGGCTTACCGGATCTGACGCGCTGGGAAAGCGCTTCCAAAACCA 1826
QY      373 yCysGluValIyLeuValProValAspGlyIyLeuGluGlyIyAspPheHisGlyProH1 393
Db      1827 CACAAAGGTGAAAAAGAAACAAAGATGGGATCGGGAACCTTGATGATCGGGCGGCCA 1886
QY      393 sValMetSerGlyTyrTrrPArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGluG1 413
Db      1887 CATATTCAATGGGTACCTCAGACACAAAGACAGACAGAGAGAGAGTGCACACCCACG 1946
QY      413 yTyrTyrCysSerGlyAspAlaIleIyLeuValAspProAlaAspProGlnIyGlyLe 433
Db      1947 CTGGCTACACACGACGACTTGGGCTTCTTG-----GATTTGACACAGTTCTCT 1994
QY      433 tMetPheAspGlyArgIleAlaAspPheIyLeuSerSerGlyValPheValSerVa 453
Db      1995 CTATGTCATGGGGAACATACATGATTTGATCAGAGCTGAGCTCGGGGAATGTCACACC 2054
QY      453 lGlyProLeu-----ArgThrArgAlaValIleuGluGlyIySerTyrValIe 469
Db      2055 GATCCCATCGAGAGCGGGTAAAGAGCGGAATCCCATC-----GTCCG 2099
QY      469 uAspValValAlaAlaProAspArgGluCysLeuGlyLeuLeuValPheProArgLe 489
Db      2100 CTATGCCATGCTGTGGGCGCAGAGCCCGCTACCTGTGCGCCCTCTC-----AC 2150
QY      489 uLeuAspCysArgAlaLeuSerSerGlyIyGlyGluAlaSerAspAlaGluValIeua1 509
Db      2151 ATTGAAGTGTCAATTAATCAGACGATCGAGAGGCTCGGAGTAC-----CTGAC 2201
QY      509 AserGluProVal-----ArgAlaTrrPheAlaAspTr 520
Db      2202 CAGTGAAGTCTGCTGCTGTGCGAGGCTTCGACCGACGCTGCTGTGTCAGACGCT 2261
QY      520 pLeu-----LysArgLeuAs 525
Db      2262 CCTATATACCGTGAACCTTGTGTCACGAGTTCATTAAACGAGGATCGAGATGTA 2321
QY      525 nArgGluAla---ThrGlyAsnAlaSerArgIleMetTrrPValGlyLeuLeuAspTrP 544
Db      2322 TTCGAGGCGGCCCAAAATGGGTCCAAGATCATTAAGTGGTC---ATCTTAACAACGA 2378
QY      544 oProSerIleAspIyGlyIyValIThrAspIyGlySerIleAsnGlnArgAlaValIe 564
Db      2379 TTTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2438
QY      564 uGlnTrrPArgSerAlaIyValAspAlaLeuTyr 575
Db      2439 CAAGATATACCAAGAGAAATCCAGAAACTCTAT 2472

```

```

REFERENCE 1 (bases 1 to 2175)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
        Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
        Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shtnsky,J.J.,
        Adame,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
        gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 2175)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
        Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
        Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shtnsky,J.J.,
        Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
        Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
        them based on alignment.
FEATURES
        Location/Qualifiers
            source
                1..2175
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                <1..>2175
                /locus_tag="HCM6513"
        gene
            Alignment Scores:
                Pred. No.: 6,47e-12 Length: 2175
                Score: 226.00 Matches: 102
                Percent Similarity: 38.15% Conservative: 75
                Best Local Similarity: 21.98% Mismatches: 213
                Query Match: 7.33% Indels: 74
                DB: 9 Gaps: 14
        US-09-750-986D-30 (1-589) x AY418305 (1-2175)
QY      181 ProAspThrIleAlaIySerpHeuPheThrSerGlySerThrIyLeuProIyAlaVal 200
Db      817 CCCAACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 876
QY      201 ProThrThrGlnArgMetLeuCysAlaAsnGlnIleMetLeuGlnThrPheProVal 220
Db      877 ATCTGAGTCAAGACATATACGTGACGACGCGACGCGTACGCGACGCGGTGACATC 936
QY      221 PheGlyGluGlu-----ProProValIleuValAspTrrPheProTrpAsnHisThrPhe 238
Db      937 CGGCCGCGCAGAAAGTCCAGACGAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 996
QY      239 GlyGlySerHisAsnIleGlyIleValIleuTyrAsnGlyGlyThrTyrTrrLeuAspAsp 258
Db      997 GCCCAGATCTACACCTGTGTGACAGGACATCCAGTGGGGGCCCAAGTTTCTTGCCGA 1056
QY      259 GlyIySProThrAlaGlnIyGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerPro 278
Db      1057 CCCGAGCCCTGAAGGGAGAGCTGTGGAACAACGCTCGG-----GAGGTGAGGCC 1107
QY      279 ThrAlaTrrLeuThrValProIySgLyTrrPgluGluIleuValGlyAlaIyLeuGluAsp 298
Db      1108 ACATCACACATGTGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1164
QY      299 SerThrLeuArgGluArgPhePheAlaArgMetIyLeuPhePhePheAlaAlaIy 318
Db      1165 GTGGCGGCTCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1224
QY      319 LeuSerGlnGlyIle----- 323
Db      1225 TTGAGACAGAAACCTCACTGCCCCGCGACGACGACCTGAAGCCCTTCAACAACAGACTGGCA 1284
QY      323 ----- 323
Db      1285 GATTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1344

```



```

Oy 245 GYlLeuValLeuTyrAenGlyGlyTyrTyrTyrLeuAapRgLyLysProThrAlaGln 264
    |||
Db 1018 AAAATGGATTTCACAGAGATATCAGGCTGCTCATGGAT----- 1059
Oy 265 GYpHeuAgluThrLeuTyrAenLeuSerGluLeuSerProThrAlaTyrLeuThrVal 284
    |||
Db 1060 -----GACCTCAAGAGGCTTCAACCCAGCTGCTTCCCCGGGTGTT 1098
Oy 285 ProLysGlyTyrGluGluLeuValGlyAlaLeuGluThrAapSerThrLeuAgluThr 304
    |||
Db 1099 CCAAGA-----CTGCTGAACCCGG 1116
Oy 305 pHeuAAlaRgMetLys-----LeuPhePhe 314
    |||
Db 1117 ATGTTTGACCGAANNMNGGACAAAGAACCAACAGCTGAAGGCTTGGACCTTT 1176
Oy 315 AlaAla-----AlaGlyLeuSerGlnGlyLe-----Trp 324
    |||
Db 1177 GCCTCCAGAGAGAAAGAGAGAGCTTCGACGGGATCATCAGAAACAAACAGCTGTGG 1236
Oy 325 AapRgLyLeu---AapRgYValAlaGluGlnHicCyGlyGluThrGlyLeuRgMetMetAla 343
    |||
Db 1237 GACCGGCTGATCTTCACAAAGTACAGTGGAGCTGGCGAAGAGTCCGGCTGATGGTG 1296
Oy 343 ----- 343
Db 1297 ACAGAGACCGCCCGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1356
Oy 344 -----GlyLeuGlyMetThrGluThrAlaProSerCyThrPheThrThrGly 359
Db 1357 CAGTTTATGAGAGATACGAGACAGACAGATGCACTCCGGGTGCTGCTGCTGCTGCTGCT 1416
Oy 360 ProLysSerMetAlaGlyTyrTrpArgAlaProGluGlnAapAlaGlnAapRgLy 379
Db 1417 GGAGACTGACCGGACGAGCATGTGGGGCCCGCATCGCGCAATTGTAATAACTGTGTT 1476
Oy 380 ProVal-----AapGlyLysLeuGluGluGlyAthrPheHicGly 391
    |||
Db 1477 GATGTGGAAAGAAATGATTAATCATGCTGCCGAGGCGGAGGCGAGGCTGTGTGAAAGGG 1536
Oy 392 ProHicValMetSerGlyTyrTrpArgAlaProGluGlnAapAlaGlnAapRgLy 411
    |||
Db 1537 CCAAAATGATTTTCAGGGCTACTGTAAGAGACCCAGCAAAACAGACAGAGCTTGGACAAA 1596
Oy 412 GYlGlyTyrTyrCySerSerGlyAapAlaLeuLeuValAapProAlaAapProGlnLys 431
    |||
Db 1597 GACGGCTGTGTTACACACAGGAGCATTTGAAATGGTTA-----CCAAATGGC 1644
Oy 432 GYLeuMetPheAapRgLyAthrGlyLeuAgluAapPheLysLeuSerGlyValAlaPheVal 451
    |||
Db 1645 ACCTTGAAATTTATGACACCGGAAAAAGCACAATTTAAGCTGGCACAGAGAAATACAT 1704
Oy 452 SerValGlyProLeuThrG-----ThrAlaValLeu 462
    |||
Db 1705 GCCCCTGAAAGATGAAATATCTACATGCGAAGTGAAGCTGTGCTGACGGTGTTC 1764
Oy 463 GYlGlyGlySer-----TyrValLeuAapValValAlaAlaProAapRgLy 479
    |||
Db 1765 CACGGGAAAGACCTGACGACCATTTCTCATTTGCAATGTGGTA-----CCAGATGTGGAG 1818
Oy 480 CyLeuGlyLeuLeuValPheProAthrLeuLeuAapCyAthrGlyLeuSerGlyLeuGly 499
    |||
Db 1819 ACATTA-----TGTTCCTGGGCCCCAAAAGAGAGA 1848
Oy 500 LysGlyAlaSerAapAlaGluValLeuAlaSerGluProValAthrAlaTrpPheAlaAap 519
    |||
Db 1849 TTGTAAGGCTGCTTTCAGAGAACTGTGACAGAAATAGAGATGTCAAAAAGCTATCTCGAA 1908
Oy 520 TrpLeuLysAthrLeuAapRgLyAla---ThicGlyAapAlaSerThrGlyLeuMetTrpVal 538
    |||
Db 1909 GATATGTGATGATGTGGAAAGATTCGTGTGAAACCATTTGAACAGGTCAAAAGCATC 1968
Oy 539 GYlLeuLeuAapThrProProSerThrAapLysGlyValThrAapLysGlySerThr 558

```

```

Db 1699 ACATTGACCCGTAATTATTTCTATCGACATGCGCTTCTGACTCCAAATGAGGCG 2028
Oy 559 AenGlnAthrAlaValLeuGlnThrArgSerAlaLysValAapAlaLeuTyr 575
    |||
Db 2029 AAAAGCCAGAGCTGCGGAATCTATTTTCAGGTGCGAGATAGATGACCTTAT 2079
RESULT 12
CR597120
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL
REMARK
1 (bases 1 to 2203)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Full-length cDNA libraries and normalization
Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue
2 (bases 1 to 2203)
Genoscope.
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
1..2203
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ004YA05"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.: 8, 83e-14 Length: 2203
Score: 244.00 Matches: 137
Percent Similarity: 38.25% Conservative: 104
Best Local Similarity: 21.75% Mismatches: 262
Query Match: 7.92% Indels: 129
Gaps: 23
US-09-750-986D-30 (1-589) x CR597120 (1-2203)
Oy 30 CyValAlaAlaArgAlaAlaAenGlyGluThrPArgArgLysSerTyrAlaGluMetPhe 49
    |||
Db 356 TGTTCAGGCTTCGGAACCAAGCAACCCATGATGATGCTTTCATATTAACAGGTTCGA 415
Oy 50 HicAenValAthrAlaAlaLeuAenSerLeuLeuProTyrGlyLeu-----SerAlaGly 67
    |||
Db 416 GAATTCGCGAGTGCATGAGCTCAGACATGATCCAGAAAGGCTTCAGACCTGCCAGAT 475
Oy 68 ArgProLeuLeuLeuValSerGlyAenAapRgLyLeuGlnHicLeuGlnLeuAapRgLyAla 87
    |||
Db 476 CAGTTTCATTCGAGATCTTCTCAAAATAGACCTGAGCTGATTAATTAACAGAGATCC 535
Oy 88 Met---TyrAlaGlyLysProTyrCyProValSerProAlaTyrSerLeuLeuSerGln 106
    |||
Db 536 TTGCTTATTCATGATGTG-----ATCGTTCCACTTATGATGATACCTTCGAAT 583

```

Db 1543 ATATCAAGCTCGTTGATGTGAGAACTGAACCTAGCGCTCGAAGGAGAGAGAGA 1602
Qy 387 lYarPheHtsGlyProHtsEvaImeserGlyTyrTPargAlaPheGlnGlnAsnIag 407
Db 1603 TATGTGTGAGAGAGAACCAATGTGTTCAAAGGCTTACTTGAAGATCCAGACAGGAGAGG 1662
Qy 407 lMaIaPheAspGlnGlnGlyTyrTyrCysSerGlyAspAlaIleYLeuAlaAspProA 427
Db 1663 AGGCGCTGAGACAGCATGCGCTTCACACTGAGACATCGGAAAATGGCTG----- 1715
Qy 427 lAspProGlnYlYSGlyLeuMetPheAspGlyArgIleAlaGluAspPheYLeuSers 447
Db 1716 -----CCGGCAGAGAACTTTAAATATATGATCGAAMAAACATATATTTAACTTGCTC 1770
Qy 447 erGlyValPheValSerValGlyPro-----LeuArgThrArgAlaVal- 461
Db 1771 AGGGAAGATATGTGTCACCCGAGAGATTGAGAACATCTACATCCGGAGCCACTGTGG 1830
Qy 462 -----LeuGlnGlyGlySer-----TyrValLeuAspValValAlaAla 475
Db 1831 CGCAATCTATGTCATGGGAGACGTTAAAGGCTTTTGGTAAAGCATTTGTGTG---- 1866
Qy 475 lAProAspArgGluCysLeuGlyLeuLeuValPheProAlGlyLeuAspCysArgAla 495
Db 1887 --CCTAACCTCGAA-----GTTATGCCCTCTCGGCGCCAGAGAGA---- 1925
Qy 495 euserGlyLeuGlyGlyGlnAlaSerAspAlaGlnValLeuAlaSerGluProValArg 515
Db 1926 -----GGAATTGAAGGAAACATATGCAGATCTCTCACAAATAGAGATCTGAAGA 1974
Qy 515 lATrPheAlaAspTrpLeuYsArgLeuAsnArgIuAla---ThrGlyAsnAlaSerA 534
Db 1975 AAGCCATTTTGGAAGATATGTGAGGTTAGAGAAAGAAAGTGAATCCATCTTTTGAGC 2034
Qy 534 rglIeMetTrpValGlyLeuLeuAspThrProProSerIleAspYlYSGlyValThrA 554
Db 2035 AGGTTTAAGCATTCATCATCTTGCATGATGTTCTCAGTTCAAAATGGCTTCTGCAC 2094
Qy 554 spYleGlySerIleAsnArgAlaValLeuGlnTrpArgSerAlaYlYValAspAla 574
Db 2095 CAACACTAAAGGTAAGAGACCTGAGCTGAGAGATCTTCAAAAACAAATAGAGAGC 2154
Qy 574 euTYr 575
Db 2155 TTAC 2159

RESULT 11
AY402444
LOCUS 2097 bp DNA linear GSS 15-DEC-2003
DEFINITION Homo sapiens PACT2 gene, VIRTUAL TRANSCRIPT, partial sequence,
AY402444
ACCESSION AY402444
VERSION AY402444.1 GI:39758430
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2097)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J.,
Adams,M.D. and Cargill,M.
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 2097)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J.,
Adams,M.D. and Cargill,M.

TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1. 2097
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2097
/gene="PACT2"
/locus_tag="NCML26"
ORIGIN
Alignment Scores:
Pred. No.: 8.18e-14 Length: 2097
Score: 244.00 Matches: 139
Percent Similarity: 36.89% Conservative: 96
Best Local Similarity: 21.82% Mismatches: 261
Query Match: 7.92% Indels: 143
DB: Gaps: 22
US-09-750-986D-30 (1-589) x AY402444 (1-2097)
Qy 30 CysValAlaAlaArgAlaAlaAsnGlyGluTrpArgArgIleSerTyrAlaGluMetPhe 49
Db 322 TGTATTAGGCTCTGGAAACAGACCAACCTATGATGCTTTCATATTAACAGTTGCA 381
Qy 50 HisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeu-----SerAlaGlu 67
Db 382 GAATTGTGGAGTGCATAGGCTCAGACATGATCAGAAAGGCTTCAAGATGCCAGAT 441
Qy 68 ArgProLeuLeuIleValSerGlyAsnAspLeuGlnHisLeuGlnIleuAlaPheGlyAla 87
Db 442 CAGTTATTTGGCATCTTTGCTCAAAATPAGACCTGAAGTGGGTATTATGACAAAGATGC 501
Qy 88 Met---TyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeuSerGln 106
Db 502 TTTCGTTATTCAGATGTG-----ATCGTTCCACTTATGATACCTTGGAAT 549
Qy 107 AspLeuAlaYlYLeuArgHisIleValGlyLeuLeuGlnProGlyIleValPheAlaAla 126
Db 550 GAA-----GCCATCAGATACATGATGCAACAAAGCTCTCTGTGTTTGTGAC 603
Qy 127 AspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValProAla--- 145
Db 604 AAGCCAGAGAGGCCAAACTTTATTAGAGGCTGTGAATAATGATTAAACGAGCCTT 663
Qy 146 -----lIePheThrArgGlyIleuAla 153
Db 664 AAATCATGATGTGTCTATGATGCTTACGGCAGTGAACCTGTGACAGAGCCAGAGGTGT 723
Qy 154 GlyArgArgThrValSerPheAspSerLeuLeuGlnGlnProGlyIleGlnAlaAsp 173
Db 724 GGGGTGAAGTCAACCATGAAGCGG---ATGGAAGA-CCTGGGAAGAGCCAAACAGACG 779
Qy 174 AsnAlaPheAlaAlaThrGlyProAspThrIleAlaYlYLeuPheLeuPheThrSerGlySer 193
Db 780 GAAGCCAAAGCTCCGCA-CCTGAAGATCTTGCAATTTGTTTCAACATGGAACCT 837
Qy 194 ThrYlYLeuProYlYAlaValProThrThrGlnArgMetLeuCysAlaAsnGlnMet 213
Db 838 ACAGGCAACCCCAAGAGACATGTCATCCAGCAAAACATATGAGCCATTTCTCAGCT 897
Qy 214 LeuLeuGlnThrPheProValPheGlyGlnGluProPro-----ValLeuValAspTrp 231
Db 898 TTGTGAAGCAACAGAGATACAGTCAATCTTGGCCAGATGATACTTGATATCTTTC 957
Qy 232 LeuProTrpAsnHisIleThrPheGlyGly-----SerHisAsnIle 244
Db 958 TTGCTCTGCGCCCATATGTTTGAAGAGATTGAGAGTGTGTAATGCTGTGTATGAGAGCT 1017

QY 67 -----GluArgProLeuLeuIleValSerGlyIleAspLeu 78
 DB 596 CAGTTATATGTCGTTTGGACAAATAATCGGCCAGAGGATCATTTG -GGAGCTGGCGTG 654
 QY 79 GluHleuGlnLeuAlaPheGlyAlaMetTyrAlaGly-IlePro---TyrCysProVa 97
 DB 655 CTACACATAT-----TTCCATGATGGTGTCCTCCCTCATATACACCTT 696
 QY 97 ILeuProAla-----TyrSerLeuSerGlnAspLeuAlaIys----- 110
 DB 697 GGGCCCTGGGGCTATCCGCTACATCATATACAGCGACATAGACACCGTATTTGGGA 756
 QY 111 -----LeuArgHisIleValGlyLeuLeuGlnProGlyIle 122
 DB 757 CAACCTCAGAAAGCTGTGCTTCTGTAGACATGTGAGAGAGAGAACCTCCAGCCCT 816
 QY 122 uValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAs 142
 DB 817 CAAGCTGATCATCTTCATGAGACCATTCAGAAAGCCCTGAAA----- 859
 QY 142 pValProAlaIlePheThrArgGlyLeuLeuAlaGlyArgThrValSerPheAspSe 162
 DB 860 -----GAGAGAGGCGAAGAGTCCGGGTGTGCTATTAACTTCACAGGC 903
 QY 162 rLeuLeuGlnGlnProGlyGlyIleGlyAlaAspAsnAlaPheAlaAlaThrGlyProAs 182
 DB 904 CGTG-----GAGGACGTGGCCACAGAGATCACAGGCTCTGTGCCCCCGACGCTGA 957
 QY 182 pThrIleAlaIysPheLeuPheThrSerGlySerThrIlySLeuProIlyAlaValProth 202
 DB 958 TCACCTCTCATTTGTGTGTTCACAGCGGACAGAGGAAACCAAAAGGTGCGATGCT 1017
 QY 202 rThrGlnArgMetLeuCysAlaAsnGlnGlnMetLeuGlnThrPhePro-----Va 220
 DB 1018 CACCCATGGGAACGTGTGCTGATTTTCAGGCTTTTCGAAAGTACAGAGAAAGTAT 1077
 QY 220 lPheGlyGlnGlnProValIleuValAspTrpLeuProTrpAsnIleThrPheGlyIle 240
 DB 1078 CTTTCGAGACAGAGACGATGTGCTCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1132
 QY 240 ySerHisAsnIleGlyIleValLeuTyr---AsnGlyGlyThrTyrTyrLeuAspAspGl 259
 DB 1133 -GAGAGAGTATTCACGCTGTGCTCATTTGCGACGAGGCGCTGTT----- 1177
 QY 259 yIysProThrAlaGlnGlyIlePheAlaGlu---ThrLeuArgAsnLeuSerGlu----- 275
 DB 1178 -----GGCTTCTTCCAGGAGATATCCGCTTCTCTCAGATACATGA 1221
 QY 276 ---IleSerProThrAlaTyrIleuThrValPro----- 285
 DB 1222 GGCTCATATGCCCATCATCTTCCCTGTGTGCCACAGACTGTGACCGATGTACAGCAA 1281
 QY 286 -----LysGlyTTP----- 288
 DB 1282 GATCTTCACGCGAAGCAACACACCATTTAAAGCGCTGCTCTCGAGTTTGCAGAAAGCG 1341
 QY 289 -----GluGlyLeuValGlyAlaLeuGlnIuArgAspSerThrLeuArgGlyIle 305
 DB 1342 TAAGCAAGCCGAGAGCTCCGAGTGAATCATCAGAAATGATATGATCTGGGATGAACCTTT 1401
 QY 305 ePheAlaArg-----MetLysLeuPhePhePheAlaAlaAla 317
 DB 1402 CTTTAAATAGATTACAGCCAGCTGTGTGTGGGTGTGCGGATGATTTGTTAACTGAGACAGC 1461
 QY 317 aGlyLeuSer-GlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHisCysGlyC 337
 DB 1462 CCAGACATACCAACAGTTCTGGGAT-----TTCCTCGGGCG 1497
 QY 337 lu-----ArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCys 355
 DB 1498 AGCTTAAAGGTGCGAGATTATGA -GGTTATGCGCAAACTGAGTGACACAGCTGAGATTA 1556
 QY 355 hPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCys 375

DB 1557 CTTTACCACTCTCTGCGCATGTGACCTTCAGGGCAGCATAGGGGCCCATCTTCCCTGCATTC 1616
 QY 375 luValLysLeuValProValAsp-----GlyLysLeuGlu 387
 DB 1617 ATATCAAGCTCGTTGATGTTTGAGAACTGAATCTACTGGGCTGCCAAAGAGAGGAGAGA 1676
 QY 387 lYArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGlnIleAsnAlaG 407
 DB 1677 TATGTGTGAGAGGACCAATATGTGTTCAAAAGGCTACTTGAAGAATTCACAGACGAGAAAG 1736
 QY 407 lAlaIleAspGlnGlnGlyTyrTyrCysSerGlyAspAlaIleLysLeuAlaAspProA 427
 DB 1737 AGGCTTCGACAGCGATGTGCTGCTTCAACCTGGAAATCGGAAATATGCTG----- 1789
 QY 427 lAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLysLeuSer 447
 DB 1790 -----CCGCGAGAACTCTTAAATATTATGATCGAAAAAGCATATATTAACTTGCTC 1844
 QY 447 eArgIlyValPheValSerValGlyPro-----LeuArgThrArgAlaVal- 461
 DB 1845 AGGAGAAATATGTGTCACCCGAGAAAGATTGAGAAATCATATCCGAGCGCAACTGTGG 1904
 QY 462 -----LeuGlnGlyGlySer-----TyrValLeuAspValValAlaAla 475
 DB 1905 CGCAATCTATGTCCATGTGGGAGCAGCTTAAAGCCTTTTGTGTAGGCACTTGTGTG----- 1960
 QY 475 lAspAspArgGlyCysLeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaAla 495
 DB 1961 --CTGACCTGAA-----GTTATGCTCTCTGGGCGCCAGAAAGAA----- 1999
 QY 495 euserGlyLeuGlyLysGlyAlaAspAspAlaGlyValLeuAlaSerGlnProValArg 515
 DB 2000 -----GSAATTTGAAGGAACATATGACATCTCTGCACAAAATAGATCTGAAGA 2048
 QY 515 lATrPheAlaAspTrpLeuLysValArgLeuAsnArgGluAla---ThrGlyAsnAlaSer 534
 DB 2049 AAGCCATTTTGGAGATATGTGTGAGGTTAGAAAGAAAGTGCATCTCATTTTGGAGC 2108
 QY 534 rGlyMetTrpValGlyLeuLeuAspTrpProProSerIleAspLysGlyGluValThr 554
 DB 2109 AGGTTAAAGCATTCATCATTCATCTTCAGACATGTTCTCAAGTTCAAAATGGCTTGTACAC 2168
 QY 554 ePlyGlySerIleAsnGlnArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaL 574
 DB 2169 CAACACTAAAGCTAAGAGACCTGAGCTGAGAGAGTACTTCAAAAACAAATAGAGAGC 2228
 QY 574 eUTY 575
 DB 2229 TTATAC 2233
 RESULT 10
 LOCUS BC026161 2518 bp mRNA linear HTC 02-MAR-2004
 DEFINITION Homo sapiens acyl-CoA synthetase long-chain family member 6, mRNA
 (CDNA clone IMAGE:4830357), containing frame-shift errors.
 ACCESSION BC026161
 VERSION BC026161.1 GI:20072873
 KEYWORDS HTC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 2518)
 REFERENCE
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheffer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, D., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,

```

Db      703 TTTATTAAGCAAGAGAGTGCATTCATCGCTTCCACAGATATGTCGTGATATCTTTC
Qy      232 LeuprotropinA1ThrhPheGlyGly-----SerHisanile 244
Db      763 TTGGCCCTCGCCCATATGTTTGAGACCGTGTAGAGTGTATATGCTGTGTGAGAGCT 822
Qy      245 GlyIleValLeuValTrpAsnGlyGlyThrTrpTrpLeuAspAspGlyLysProThrAlaGln 264
Db      823 AAGATATGAGATTTTCCAGAGATATATCAGCGCTGCTATATGAC-----864
Qy      265 GlyPheAlaGluThrLeuArgAsnLeuSerGluLysSerProThrAlaTrpLeuThrVal 284
Db      865 -----GACCTCAAGGCTCTTCAAGCCCACTCTTCCCTCGTGCTT 903
Qy      285 ProlyseGlyTrpGluGluLeuValGlyAlaLeuGluArgAspSerThrLeuArgGlyLys 304
Db      904 CCCAGG-----CTGCTGAACCGG 921
Qy      305 PhePheAlaArg-----MetyLeuPhePhePheAlaAla-----317
Db      922 ATGTTGACAGATGACAAACAAACCTTCTTGAAGCATGAGCTGTGACCTTTCCTTC 981
Qy      318 -----GlyLeuSerGln-GlyIle-----TrpAspArg 326
Db      982 CAAAGGAAAGAGCGCGAGCTTGCAGTGCATCGTCAGAAACACAGCGCTGGATTA 1041
Qy      326 GLeu--AspArgValAlaGluGlnHisCysGlyGlyLysArgMetMet-----342
Db      1042 ACTCATCTTCCACAAAGATACAGTCGAGCTCGGTGGGAAAGTCGGCTGATATCACAG 1101
Qy      342 -----342
Db      1102 AGCAGCCCGGTGCTGACAGTGTGACGTTTCTGAGACAGCGCTGGCTGCCAGTT 1161
Qy      343 ----AlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProle 361
Db      1162 CTATGAGAGCTACGACAGACGAGTGCATCGTGTGTGCTGTGCTGACGCTGCCGAG 1221
Qy      361 uSerMetAlaGlyTrpIleGlyLeuProAlaProGlyCysGluValLysLeuValProVa 381
Db      1222 CTGACGCGAGCGCATGTTGAGCCCATGCTTCATTAATTAAGCTTGTGATGATG 1281
Qy      381 LAsp-----GlyLeuGluGlyValArgPheHisGlyProH 393
Db      1282 GGAAGAAATGATTAATTCCTGCGCATCAAGGCGAGGCTGAGTGTGTGAAAGGCGCA 1341
Qy      393 eValMetSerGlyTrpArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGly 413
Db      1342 TGTGTTCAAGGCTATCTTAAGAGACCAAGACAGACAGCTGAGCCCTGATTAAGATG 1401
Qy      413 yTrpTrpCysSerGlyAspAlaIleLysLeuAlaAspProAlaAspProGlnLysGly 433
Db      1402 CTGCTTACACACGCGGAGCATTTGGAAATGCGCTG-----CCAAATGCGACCTT 1449
Qy      433 uSerPheAspGlyArgIleAlaGluAspPheLysSerSerGlyValPheValSerVa 453
Db      1450 AAGAGATTATGACAGAGAAAAACATATTTAACTAGCCCAAGAGAGATACATAGAC 1509
Qy      453 LglPro-----LeuArgThrArgAlaVal-----LeuGluGly 464
Db      1510 AGAAAGATTTGAATATCTACTGCGAGTGAAGCGGTGCGCCAGGTGTTTTCACAG 1569
Qy      464 yGlySer-----TyValLeuAspValValAlaAlaProAspArgLysGly 481
Db      1570 AGAAAGCTTCAAGCGCTTCTCTATACAGTGTGAGTA-----CCGACGCTTGAAG 1623
Qy      481 uGlyLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGly 501
Db      1624 A-----CCGTCCTGGGACAGAAAGAGGCTTTCAGAGG-----1657
Qy      501 uLaserAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 521

```

```

Db      1658 ----TCCTCGAAGAACTGTGACGAAACAGAGATATCATTAAGCTATCTCGACGACTT 1713
Qy      521 uLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerAlaGlyMetTrpValGlyLeu 541
Db      1714 GTTGAACCTTGGAGAGAAAGCC---GGTCTGAAGCATTGAAACAGCTCAAGGCAATTGC 1770
Qy      541 uAspThrProPro-----SerIleAspLysGlyGluValThrAspLysGlySerIleAs 559
Db      1771 TGTGACCCCGGAATTAATTTCTATTTCACAAAGCGCTTGTGATCTCAACAGTGAAGCG 1830
Qy      559 nGlnArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaLeuTrp 575
Db      1831 GAGGCGAGAGTACGGAATTAATTCAGTGTGCGACAGATAGTAACCTGTAC 1879

RESULT 9
LOCUS      CR606980      2548 bp      mRNA      linear      HTC 21-JUL-2004
DEFINITION      Full-length cDNA clone CS0DJ003YJ20 of T cells (Jurkat cell line)
ACCESSION      CR606980
VERSION        CR606980.1 GI:50487787
KEYWORDS       HTC; CNSL1_cDNA.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
REFERENCE      Li,W.B., Gruber,C., Jessee,J. and Polyses,D.
AUTHORS        Full-length cDNA libraries and normalization
JOURNAL        Unpublished
REMARK         Contact : Feng Liang Email : fliang@lifetech.com URL :
                http://fulllength.invitrogen.com/InvitrogenCorporation 1600
                Faraday Avenue
                2 (bases 1 to 2548)
                Genoscope.
                Direct Submission
                Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
                BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
                - Web : www.genoscope.cns.fr)
COMMENT        1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
                end enriched, double-strand cDNA was digested with Not I and cloned
                into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
                was normalized. Library was constructed by Life Technologies, a
                division of Invitrogen.
FEATURES       Location/Qualifiers
                source          1..2548
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CS0DJ003YJ20"
                /issue_type="T cells (Jurkat cell line) Cot
                10-normalized"
                /plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.:      1 286-14      Length:      2548
Score:          253.00      Matches:      148
Percent Similarity: 38.10%      Conservative: 97
Best Local Similarity: 23.02%      Mismatches: 246
Query Match:    8.21%      Indels:      154
DB:             3      Gaps:      27
US-09-750-986D-30 (1-589) x CR606980 (1-2548)
Qy      30 CysValAlaAlaArgAlaAlaAsnGlyLysTrpArgArgIleSerTrpAlaGluMetPhe 49
Db      476 TGTCTTGTTTCAAGAGAGCTTAAGACGCTTACACAGTGTGCTTCAACAGAGAGTGCC 535
Qy      50 HisAsnValArgAlaIleAlaGlnSerLeuLeuProTrpGlyLeuSerAla-----66
Db      536 GACAGGCTGAATTTCTGTGGGTCCGAGCTTTCACAGACAAATTGTAACATGCACTGAT 595

```

Db 1382 ACAGAGCAGACCCGGTCTGCCACAGCTGACGTTTCTGAGGACAGGCTGGCTGC 1441
Qy 343 -----AlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrGly 359
Db 1442 CAGTTTATGAGAGCTACGACGACGAGGATGACGCTGGTCTGGCCCTGACCTGGCCC 1501
Qy 360 ProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGluValIlyLeuVal 379
Db 1502 GGAGACTGACGCGACGCGCATGTTGGAGCCCCCATGCTTGCATATTAATGAAAGCTTGTG 1561
Qy 380 ProValAsp-----GlyLeuGluGluGlyAArgPheHisGly 391
Db 1562 GATGTGGAGAAATGAATTAATCTGGCATCCAGCGGAGGATGAGTGTGTGAAAGGG 1621
Qy 392 ProHisValMetSerGlyTyrTrpArgAlaProGluGlnHisAlaGlnAlaPheAspGlu 411
Db 1622 GCAATGTGTCTTCAAGGCTTACCTTGAAGAGCCAGCAAGAACAGCTGAAGCCCTGGATAA 1681
Qy 412 GluGlyTyrTyrCysSerGlyAspAlaIleLeuLeuAlaAspProAlaAspProGlnLys 431
Db 1682 GATGCTGCTGATACACAGCGGGGACATTGGAAATGCGTG-----CCAAATGGC 1729
Qy 432 GlyLeuMetPheAspGlyArgIleAlaGluAspPheLeuSerSerGlyValIleVal 451
Db 1730 ACCTTGAAGATTATGACAGAGAAACACATATTAACTACCCAGAGAGATACATA 1789
Qy 452 SerValGlyPro-----LeuArgThrArgAlaVal-----Leu 462
Db 1790 GCACCGAGAAAGATTGAAATATCTACTCTCGGAGTGAAGCCCTGGCCCGAGCTTTGTC 1849
Qy 463 GluGlyGlySer-----TyrValLeuAspValValAlaIleAlaProAspArgGlu 479
Db 1850 CACGGAGAAAGCTTTCAGGCGCTTCTCATAGCAGTTGTGGTA-----CCGACGTTGAG 1903
Qy 480 CysLeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGly 499
Db 1904 AGCCTA-----CCGTCTCGGACACAGAGAGAGGCTTTACAAAGG-----1942
Qy 500 LysGluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAsp 519
Db 1943 -----TCTTCGAAAGACCTGTCAGAGAACAGATATCAATAAGCTATCTCGAC 1993
Qy 520 TrpLeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGly 539
Db 1994 GACTTGTTGAACCTTGGAGAGAGCC---GGTGAAGCATTGTAACAGTCAAGG 2050
Qy 540 LeuLeuAspThrProPro-----SerIleAspLysGlyGluValThrAspLysGlySer 557
Db 2051 ATTGCTGTCCACCCGGAATTAATTTCTATTGACACAGGCTTCTGACTCCACACCTGAAG 2110
Qy 558 IleAsnGlnArgAlaValLeuGlnTrpArgSerAlaLeuValAspAlaLeuTyr 575
Db 2111 GCGAAGAGCCACAGCTACGGAACATAATTTCAGTCCGACATGATGAACGTGAC 2164

RESULT 8
AY402446 1897 bp DNA linear GSS 15-DEC-2003
LOCUS Mus musculus FACL2 gene, VIRTUHL TRANSCRIPT, partial sequence.
DEFINITION genomic survey sequence.
ACCESSION AY402446
VERSION AY402446.1 GI:39758432
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 1897)
Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejaritwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Interfering nomenclature evolution from human-chimp-mouse orthologous
gene titos

JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1897)
AUTHORS Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejaritwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1. 1897
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1. >1897
/gene="FACL2"
/locus_tag="HCM1226"
ORIGIN
Alignment Scores:
Pred. No.: 8.11e-15 Length: 1897
Score: 253.00 Matches: 154
Percent Similarity: 36.89% Conservative: 81
Best Local Similarity: 24.18% Mismatches: 259
Query Match: 8.21% Indels: 143
DB: 9 Gaps: 27
US-09-750-986d-30 (1-589) x AY402446 (1-1897)
Qy 30 CysValAlaAlaArgAlaIleAlaAsnGlyGluTrpArgArgIleSerTyrAlaGluMetPhe 49
Db 127 TGTATTGATTTCTGAGAGCCAAACAGCCCTTATGATGATTTCTTCAAAAGAGGTGCA 186
Qy 50 HisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeuSer-----AlaGlu 67
Db 187 GAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 246
Qy 68 ArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeuAlaPheGlyAla 87
Db 247 CAGTTTATGAGAGCTTCTTCTTCAAAACAGCCGAGTGGGTGATCTCGACAGAGATGC 306
Qy 88 Met-----TyrAlaGlyLeuProTyrCysProValSerProAlaTyrSerLeuLeuSerGln 106
Db 307 TTTCTTACTCATATGATG-----GTGCTCCGCTCTATACACCTTGGAGCT 354
Qy 107 AspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaIle 126
Db 355 GAC-----GCCATCACCTACATATGAAACAAAGCTGAATCTCTGATTTTGTCTGAC 408
Qy 127 AspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValProAla--- 145
Db 409 AAGCCAGAAAAGCCAAACTTTATTGAAAGGTGTAAGAAACAAAGTTAACACCATGCTT 468
Qy 146 -----IlePheThrArgGlyGluLeuAla 153
Db 469 AAAATATATGATCATCTGACATCTTACGCGAGTGAATCTGTTGAGACAGAGCAAGAGTGT 528
Qy 154 GlyArgArgThrValSerPheAspSerLeuLeuGluGlnProGlyGlyIleGluAlaAsp 173
Db 529 GGGGTGAATATCATCATGCTCAAAGCT---CTGAGAGACCTTGGAGAGAGGACAGAGTG 585
Qy 174 AsnAlaPheAlaIleAlaThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySer 193
Db 586 AAGCCC---AAGCTTCCAGAACCCGAAAGATCTTGCATATTTGTTTCAACATGGAACCT 642
Qy 194 ThrLysLeuProLys---AlaValProThrThrGlnArgMetLeuCysAlaAsnGln 212
Db 643 ACAGGACACCCCAAGAGAGCATGATCATCTACCAAAACATTTTAAACAGCTGCTCAGGT 702
Qy 213 MetLeuLeuGlnThrPheProValPhe---GlyGluGluProProValIleuValAspTrp 231


```

Db 1733 ACTGCGGGCGCGAGAACGTCGCCGCTGCTATCGAGAGAGCCGTGAAGATGAG--- 1789
Qy 465 GlysertyrValleuaspValValAlAlA-----ProAspArgGluCys 480
Db 1790 -----CTGCCCATCATCAGACGTCCTGATCGGGAGACAGAGAGATTC 1837
Qy 481 LeuGly----LeuValPheProArgLeuLeuasp----- 491
Db 1838 CTGTCCATGCTGCTTACCTGAAAGTGAACGCTGACCAAGAGAGCTTGAGCTTACAGAC 1897
Qy 492 -----CyAspAlaLeuSerGlyLeuGlyValAla 502
Db 1898 AGCCTGACCGAGAACGCTGAGATTCCTGACGAGGTGGGACGACAGCAGACCGTC 1957
Qy 503 SerAspAlaGluValleuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuAsp 522
Db 1958 TCCGAGATCGTGGGCGAGAGACGAGCGCTGTACACGAGCCATTCACGAGGATTCAG 2017
Qy 523 ArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAsp 542
Db 2018 AGGATGAAACCGAACCGCCGACGCCCTTACCACATCCAGAGTGGGCTTCTTCAA 2077
Qy 543 ThrProSerIleAspIleValGlyValThrAspIleSerIleAsnArgAla 562
Db 2078 CGTACCTTCTCATTTGAGGTGAGAGCTGGGACCCATGAATGAACGAGCTCAGC 2137
Qy 563 ValLeuGlnTrpArgSerAlaValAlaAspAlaLeuTrpArgGlyLeuAspGln 580
Db 2138 GTTCTGAGAGAGTAAAGATATCATCATCTTTTATCATAGACAAAGAACAG 2191

RESULT 6
AY109410 2505 bp mRNA linear HTC 17-OCT-2002
LOCUS AY109410
DEFINITION Zea mays CL741.1 mRNA sequence.
ACCESSION AY109410
VERSION AY109410.1 GI:21213124
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 2505)
Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
Maize Mapping Project/Dupont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 2505)
Coe, E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schubler, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.
Location/Qualifiers
1. 2505
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_id="Maize Mapping Project/Dupont Consensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed Dupont contigs; this resource was
assembled by Dupont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize

```

```

ORIGIN Mapping Project"
Alignment Scores: 7.91e-16 Length: 2505
Pred. No.: 264.50 Matches: 154
Score: 36.62 Conservative: 110
Best Local Similarity: 21.368 Mismatches: 220
Query Match: 8.588 Indels: 237
DB: 3 Gaps: 37
US-09-750-986d-30 (1-589) x AY109410 (1-2505)
Qy 23 LysThrArgProGluGlnThrCysValAlaAlaArgAlaAlaSerGlyGluTrpArg 42
Db 425 AGAATCGCCCAACGAGGACAGTT-----GAGACTCAAGTGG 463
Qy 43 IleSerTyraGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuProTy 62
Db 464 ATGACGTATGAGAAAGCTAGACAGACAGAACTGCAATAGTTCTGTATTATTCAT 523
Qy 63 GlyLeuSer-----AlaGluArgProLeuIleVal 73
Db 524 GGAATATCTGAAGTGCACGAAATGCTATATTATTAACAGACCGGATGATCAT 583
Qy 74 SerGlyAsnAspLeuGlnIleGlnLeuAlaPheGlyValMetTyraGlyIlePro 93
Db 584 -----CTTACCATGCTTGTGGCGCATCTCA-----TATGATCTGTG--- 622
Qy 94 TyrCysProValSerProAlaTyrSerLeuLeuSerGlnAspLeuAlaLysLeuArgHis 113
Db 623 -----CCGCTTATGATCTCTGCGCCAGAT-----GCAGTCAAGTTT 661
Qy 114 IleValGlyLeuGlnProGlyLeuValPheAlaAlaAspAlaAlaProPheGlnArg 133
Db 662 ATAGTAAACCAACGCAACGATGAGATTAATCTGTGTG-----CCCA 706
Qy 134 AlaIleGluThrIleLeuPro----- 140
Db 707 ACTCTAAGCCCTATTTTAAAGCTTTATTACTCAATGCCGTGTTGCTNNNNNNNNNN 766
Qy 141 -----AspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyArg 155
Db 767 NNNNNNNNNNNNTTCAATATATGATCATCTGACACGACCAACTCTGAATGCC----- 817
Qy 156 ArgThrValSerPheAspSerLeuLeuGlnProGlyGlyIleGluAlaAspAsnAla 175
Db 818 AAATATATGACGTACACAGGCTACCAATGAG-----GAAAGATGATCTCAAACT 871
Qy 176 PheAlaAlaThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLys 195
Db 872 TTTCGTCCACCAAACTGTAAGATATAGCCATATCTGCTACCCAGTGACATCTGGC 931
Qy 196 LeuProLysAlaValProThrThrGlnArgMetLeuCysAlaAsn-----GlnGln 212
Db 932 ACACAAAGGAGGAGCTGACTTCTCATGAGAACTATATGCAATAGAGAGGCTCAAGT 991
Qy 213 MetLeuGlnThrPheProValPheGlyGluGlnProProValLeuValAspTrpLeu 232
Db 992 TTAGGATTAAGTTTACCC-----TCTGATGTGTACATCTGTTCTA 1036
Qy 233 ProTrpAsnHisThrPheGlyGlySerHisAsnIleGlyIleValLeuTyraGly--- 251
Db 1037 CTTTGGCTCACATCTATGAGAAAGCTAACAGAGTTCATCTTCACTATGTTGCTGCC 1096
Qy 252 ---GlyThrTyraLeuAspArgGlyLysProThrAlaGlnGlyPheAlaGluThrLeu 270
Db 1097 ATTGATTTTCAACAGGAGTAACTGAAG-----CTGATG 1132
Qy 271 ArgAsnLeuSerGluLeuSerProThrAlaTyraLeuThrValPro-----Lys 286
Db 1133 GATGATTTGCTGTTGAGAACCAACAAATATTCGAAGTGTCCCGGTTATATTAACAA 1192

```

CDS

```

/tissue_type="ovoiduct"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="2 days pregnant adult"
29. 2194
/note="unnamed protein product: lipidosis-related protein
lipidosis (LocustLink|94180, GB|NM_051178, evidence:
BLASTN, 99%, match=2587)
putative"
/codon_start=1
/protein_id="BAC35657.1"
/db_xref="GI:26344001"
/translation="MPRGSAHYCCSLSRDSNMPDSDQOQASLTSQNSQSLI
DQDTLSKSPSHGLSAPKAPASLDGAELWTRADGRVLRLEPCTPRPVY
HGFYEALDRYGNLSALGFRKDKMERISYYQYLARKYAKFLKGLERASVAL
HSPSEWFFADVGVFAGIVTGITTSSEACQYISDGRANVIVDTQKOLELXK
IMDPLKAVVLYOEPBPCKMANVYMEELIELGQVPEBALDAIIDQOQKCCVL
VYSGTGNKQKVMLSQDNTWTARYSQADIDPAVQOEVVSYIPLSHIAQITD
LMTGIQAOVCFADPDALKGTIVNTUREVEPVSHPVPRWVEKIMERIEVAQSGE
IRRMILMAMSVTLBQNLTCPSNDLKPFTRADYLVAVRQALGAKQKMFYGA
PMTAEIQRFLGLNIRLYAGYGLSESTGPHFMSPVYRYSGRVPGCRVLYND
ADGIEICLNGRTIFMGYLNMDKTCALIDSEGLHTGMDGLDADGFLYITGLKEL
ITTAGENVPVPIBEAVKMEPLIISAMLIIGORKFLSMILTLTKTLDETSRPNDS
LTEQAVECCORNSKASTVSEIYQGRBEATYQALHEGIQVNNANARPHIQKMAIL
QRFDSISGELGPTMKLRLVLEKTIIDITFQEQKQ"
2678..2683
polysignal
/note="putative"
2698
polysite
/note="putative"
2698
ORIGIN
Alignment Scores:
Pred. No.: 1,87e-16 Length: 2698
Score: 271.00 Matches: 154
Percent Similarity: 35.55% Conservative: 87
Best Local Similarity: 22.71% Mismatches: 287
Query Match: 8.79% Indels: 150
Gaps: 24
US-09-750-986d-30 (1-589) x AK054103 (1-2698)
QY 12 G|YArgIleLeuGluAValGluGlnHisThrAlaIleThrArgPro-----GluGln 28
Db |||||::: |||||::: |||||
QY 281 GGAAGAGTGGCGCTGCGCTGGAACCTTCTGTACTGACGAGTCCCTTATACCGTGACAG 340
Db |||||::: |||||::: |||||
QY 29 ThrCysValAlaAlaArgAlaAlaAsnGly----- 38
Db |||||
QY 341 ATGTTCTACAGAGCGCTTACAGTAAGTACGGAACCTCAGTCTCGGGCTTCAAGCGCAG 400
Db |||||
QY 39 ---GluThrArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGln 57
Db ::||| |||||::: |||||::: |||||
QY 401 GACAAAGTGGAGCGTATCTTACTTACTACAGTACTACTGATGGACGCAAAAGTGGCAAA 460
Db |||||
QY 58 SerLeuLeuProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAsp 77
Db |||||
QY 461 GGCCTTCTTGAAGCTCGGCTTACGAGCGTCCACAGCGTGGCGATCCTTGGCTTCAACTCT 520
Db |||||
QY 78 LeuGlnHisLeuGlnLeuAlaPheGlyAlaMetTyrAlaGly-----IlePro 93
Db |||||
QY 521 CCAGAAATGCTTCTTCTGTGCAAGTGGCAGACAGTGTTCGACAGGGGCAATTCACAGCGATC 580
Db |||||
QY 94 TyrCysProValSerPro---AlaTyrSerLeuLeuSerGlnAspLeuAlaIleValLeuArg 112
Db |||||
QY 581 TACACCAACCAAGTCCCGGAGAGCTTGCACATCACTCATGACAGTCCGA----- 631
Db |||||
QY 113 HisIleValIleValLeuLeuGlnProGlyLeuValPheAlaIleAspAlaIleProPheGln 132
Db |||||
QY 632 -----GCCAATGTCAATCGGTGACACA-----CAG 658
Db |||||
QY 133 ArgAlaIleGluThrIleLeu----- 139
Db |||||
QY 659 AAGCAGCTGAAAGATCTCTGAAGATCTGAAAGACTTGCACACTTAAAGGACAGTGGTG 718
Db |||||
QY 140 -----ProAspAspValProAlaIlePheThrArgIleGluLeuAla 153

```

```

Db 719 ATATACCAAGAACCCCTCCAAAGAAAGATGCTTAACTGTACAGATGAGAGGACTATA 778
QY 154 G|YArgArgThrValSerPheAspSerLeuLeuGlnProGlyGlyIleGluAlaAsp 173
Db 779 GAA-----CTGGGCGCAAGAAAGTGGCTTGAAGAGCCCTTA 811
QY 174 AsnAlaPheAlaAlaThr---GlyProAspThrIleAlaIleAspLeuPheThrSerGly 192
Db ::||| |||||
QY 812 GATGCCATATCATGATACCCAGAACCCAAACAGAGCGTGTGCTGCTCATCATGAGGC 871
Db 779 GAA-----CTGGGCGCAAGAAAGTGGCTTGAAGAGCCCTTA 811
QY 193 SerThrIleLeuProIleValAlaValProThrThrGlnArgMetLeuCys----- 208
Db ::||| |||||
QY 872 ACCACGGGAAACCCCAAGGGGTGATGCTGATGCAAGACAATATCATATGACAGACGCG 931
QY 209 ---AlaAsnGlnIleMetLeuLeuGlnThrPheProValPheGlyGluIleProProVal 227
Db ::||| |||||
QY 932 TATGGCAGTCAAGCTGGGACATCCACCCAGCAAGATG-----CAGCAGAGAGTA 982
QY 228 LeuValAspTrpLeuProTrpAsnHisThrPheGlyGlySerHisAsnIleGlyIleVal 247
Db |||||
QY 983 GTGGTACGCTACTTGGCTCTCAGCCACATGCGTCCAGATCTTACAGACCTGTGGACAGGC 1042
QY 248 LeuTyrAsnGlyGlyThrTyrTyrLeuAspArgGlyLysProThrAlaGlnGlyPheAla 267
Db ::||| |||||
QY 1043 ATCCAGTGGGAGCCCGCAGTCTGCTTGGCAGATCCGAGCCCGTGAAGGGAGCCCTGTGA 1102
QY 268 GluThrLeuArgAsnLeuSerGluIleSerProThrAlaThrIleuThrValProIleGly 287
Db |||||
QY 1103 AACACACTGGCG-----GAGGTGAGCCCAATCCACATCGGAGTGGCGGTGTG 1153
QY 288 TrpGluGluLeuValGlyAlaLeuGluArgAspSerThrLeuArgGluArgPhePheAla 307
Db |||||
QY 1154 TGGAGAAATCATATGAAAGATCCAG--GAGGTGGCGCTCATGCTGCTTCATCCGC 1210
QY 308 ArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSerGlnGlyLe----- 323
Db |||||
QY 1211 CGCAGATGCTGCTGGGCGCATGTCAGTCACTTGAACAGAACCTTACCTGCTCCAGC 1270
QY 324 -----TrpAspArgLeuAspArgValAla 331
Db |||||
QY 1271 AATGACCTGAAGCCCTTACACAGACAGTGGGAGATTACTAGTGTAGCCAGAGTCCGT 1330
QY 332 GluGln----- 333
Db |||||
QY 1331 CAGGCTCTGGCGCTTGGCAAGTGTCAAGAAACTTCTTACGAGACAGCTCCGATGACGCA 1390
QY 334 -----HisCysGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThr 348
Db |||||
QY 1391 GAGACGAGCGCTTCTTCTGAGGCTCAACATCCGCTGTATGACAGCTTATGCGCTTACAGC 1450
QY 349 Glu---ThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyr--- 366
Db |||||
QY 1451 GAGAGCAGAGCCCGCCAC-----TTCATGTCCAGCCCTTCAACAATATCGGCTTACAGC 1504
QY 367 IleGlyLeuProAlaProGlyCysGluValIleLeuVal-----ProValAspGlyLys 384
Db |||||
QY 1505 TCCGGCAGAGGTGTATCCCGGCTCCCGGTGAACCTGTGAATCAGATGTGCAATGGCATC 1564
QY 385 LeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGluGln 404
Db |||||
QY 1565 GGTGAGATGTGCTGAGGCGCCGACACCATCTTATGGGTATATGAAATGAGAGACAAA 1624
QY 405 AsnAlaGluAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleIleValLeuAla 424
Db |||||
QY 1625 ACCTGTAGGCGCATTAATATGGAAGGTGGCTGACACAGGTATGAGTGGCGCCCTG--- 1681
QY 425 AspProAlaAspProGluIleGlyLeuMetPheAspGlyArgIleAlaGluAspPheLys 444
Db |||||
QY 1682 -----GATCTGACGCGCTTCTCATACACCGGGGCTCAAGAAATTAATCATC 1732
QY 445 LeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValIleGluGly 464
Db |||||

```


FEATURES
source

Location/Qualifiers
1. 544
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"
/strain="USD110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/note="Vector: pIndigoS36; Site_1: HindIII"

ORIGIN

Alignment Scores:

Pred. No.: 1,06e-20 Length: 544
Score: 301.50 Matches: 71
Percent Similarity: 59.64% Conservative: 28
Best Local Similarity: 42.77% Mismatches: 52
Query Match: 9.78% Indels: 15
DB: 8 Gaps: 3

US-09-750-986d-30 (1-589) x AZ935488 (1-544)

QY 425 AAGPProAlaAapProGlnLysGlyLeuMetPheAapGlyArgIleAlaGluAapPheLys 444
Db 2 GATAGCGCGCGCGCGCGCGCGCTGTTCTTCGACGCGCGTGTCCGCGAGGAGCACTTCAAG 61
QY 445 LeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlnGly 464
Db 62 CTCATATTCGGGACCTGCTGACGCTGCGACCGCTGCTGCGCGCGTGTCCGCGAGGAGCACTTCAAG 61
QY 465 GlySerTyValLeuAapValAlaValAlaAProAapArgGluCysLeuGlyLeu 484
Db 122 GCGCGCGCGCGCGAGATATCTGCTGCTGCGCGTATCGCGCGCGAGCGCGCTTCTG 181
QY 485 ValPheProArgLeuLeuAapCysArgAlaLeuSerGlyLeuGlyValLeuAap 504
Db 182 GTGTTCCCGGAACTCTGCGCGCGCGCGCATCCCGCTTGGCGGAGCGCGCATGTG 241
QY 505 AlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAapTrpLeuLysArgLeu 524
Db 242 AATGATGTGCTGCGACATGCAAGATAGGGGCGCGATCTCTCAGGGGCTTGGCAAGCTG 301
QY 525 AapArgGlu-----AlaThrGlyAapAlaSerArgIleMetTrpValGlyLeu 541
Db 302 AAGCAGAGGCGCGCAATCTCTCCGCGAGCGCGCGCGCTG-----CTGCTC 352
QY 542 AapTrpProProSerIleAapLysGlyValValThrAapLysGlySerIleAapGlnArg 561
Db 353 GCCGAAACCGCATCGTCTGCGCGCGCGAGATCACCGCAAGGCGCTACATCAACGCGCG 412
QY 562 AlaValLeuGlnTrpArgSerAlaLysValAlaPheLysArgGlyValAapGlnSer 581
Db 413 GCGGTCTGACGCGCGCGCGAGAGCGCGT-----GCG 445
QY 582 MetLeuArgAapGlnLys 587
Db 446 CCGCTGAATGACGATCGG 463

RESULT 5
AK054103 2698 bp mRNA linear HTC 03-Apr-2004
LOCUS AK054103.1 GI:26344000
DEFINITION Mus musculus 2 days pregnant adult female oviduct cDNA, RIKEN
product:lipidosis-related protein lipidosis, full insert sequence.
ACCESSION AK054103
VERSION AK054103.1 GI:26344000
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2

TITLE Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20493374
PUBMED 11042159

REFERENCE

AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN Integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multichannel sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2698)

REFERENCE

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, Y., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (16-Jul-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/

FEATURES

source

Location/Qualifiers
1. 2698
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:E230019G03"
/db_xref="taxon:10090"
/clone="E230019G03"
/sex="female"


```

Db      243 ATGTATCGGCGGCGGCTGTCGGAACAATCGCCGCTGTGCACCGGTCACATTTCTC 184
Qy      360 oLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCyseGluValLysLeuValPr 380
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      183 GCGGAGCTCCCGGCAATATCGCGCTCCGATACCGGCTCGAGCTGAAGCTCGCAC 124
Qy      380 oValAspGlyLysLeuGluGlyValArgPheHisGlyProHisValMetSerGlyTyrTrp 400
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      123 CTCGCGGCAACAAGCTGAAGTGCAGCGCTGCTGCCCAAGCTCACGCTGGGTTACTGGAA 64
Qy      400 gAlaProGluGlnAsnAlaGlnAlaPheArgGluGluGlyTyrTyrCyseSerGlyAspAl 420
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      63 GCGCGCGAAGCTGACCAAGGAGCGTTCGTGCAGGAGGCGCTTTATCTCTCGTGATGC 4
Qy      420 a 420
        |
Db      3 G 3

RESULT 3
AZ933744/c 714 bp DNA linear GSS 24-APR-2001
LOCUS      B1_Ba0001A14r B. japonicum BAC library Bradyrhizobium japonicum
DEFINITION genomic, genomic survey sequence.
ACCESSION  A2933744.1 GI:13775804
VERSION     A2933744
KEYWORDS   Bradyrhizobium japonicum
SOURCE      Bradyrhizobium japonicum
ORGANISM   Bradyrhizobium japonicum
REFERENCE   1 (bases 1 to 714)
AUTHORS     Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
            Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
            A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome Res. 11 (8), 1434-1440 (2001)

TITLE       genome Res. 11 (8), 1434-1440 (2001)

JOURNAL     genome Res. 11 (8), 1434-1440 (2001)
MEDLINE     21376150
PUBMED      11483585
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: twing@clemson.edu
            Class: BAC ends
            High quality sequence start: 23
            High quality sequence stop: 664.
            Location/Qualifiers
                1..714
                /organism="Bradyrhizobium japonicum"
                /mol_type="genomic DNA"
                /strain="USDAL10"
                /db_xref="taxon:375"
                /lab_host="E. coli"
                /clone_lib="B. japonicum BAC library"
                /note="Vector: pIndigo56; Site_1: HindIII"

ORIGIN
Alignment Scores:
Pred. No.:      6.33e-22      Length:      714
Score:          315.00      Matches:      77
Percent Similarity: 52.23%      Conservative: 40
Best Local Similarity: 34.38%      Mismatches: 105
Query Match:    10.22%      Indels:      3
DB:             8      Gaps:      1

US-09-750-986d-30 (1-589) x AZ933744 (1-714)

Qy      202 Th-ThrGlnArgMetLeuCySAIAaenGlnMetLeuLen-----ThrPhePro 219
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      703 ACACCGAGGAGCATGATGACTCGAGCCAGCGCAAGCGCAAGCGCATGTCGACATTCCTTG 644

```

```

Qy      220 ValPheGlyGluGluProProValLeuValAspTyrLeuProTrrPasnHisThrPheGly 229
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      643 GACATGAGCTGAGCAGCATCTCGATTCGATCGAGCTGCGCTGATCGCTCATCATTCGAGC 584
Qy      240 GlySerHisAsnIleGlyIleValIleValIleValIleValIleValIleValIleVal 259
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      583 TCCAACCCACATTTCAACTGAGTGAAGCCGATCGGCGGCTCATCTATATCGACGGCGGC 524
Qy      260 LysProThrAlaGlnIlePheAlaGluThrLeuArgAsnLysSerGluLysLeuProThr 279
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      523 GAGCTTCGACCGCGCTCTTTCGAGCTGCGCTCGCAACTATACAGCGGATCGGACG 464
Qy      280 AlaTyrLeuThrValProLysGlyTyrGluGluLeuValGlyAlaLeuGluArgAspSer 299
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      463 GTCTATTTCACAGCTGCGCGCGCTTCCACATGCTGATCGCGTCATTCGCTCGTGACGA 404
Qy      300 ThrLeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaGlyLeu 319
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      403 GCGTTCGTCGCGCTTCCTTCACGAGGTGAAGTTCGCTTTCACGCGGCTCAGCGCTG 344
Qy      320 SerGlnGlyIleTrrAspArgLysLeuArgValAlaGluGlnHisCyseGlyLysGlyLe 339
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      343 CCGCAAGACCTGTGAACGCGCTCGTACATTCCTCGTCAATACCGTCCGCGCGCTG 284
Qy      340 ArgMetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCyThrPheThrGly 359
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      283 CCGATGTATCGCGCTCGGCGCTCGAGGAACCTCGCGCTGCGACCGCTCGCATTTTC 224
Qy      360 ProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCyseGluValLysLeuVal 379
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      223 CTCGCGGAGCGCTCCGTAATATCGGCGGCTGACATACCGGTACCGAGCTGAAGCTGTC 164
Qy      380 ProValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTrp 399
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      163 ACCTCGGCGCCAGCTGATAGTCCGCGGCGGCCACCGCTCTCGCAGGTTATGCA 104
Qy      400 ArgAlaProGluGlnAsnAlaGlnAlaPheArgGluGlyTyrTyrCyseSerGlyAsp 419
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      103 ACAGCTCGTG-CTGCGCAGAGGAGATAGCCGACCATGCGCTTTATCTCATCGTGGGT 45
Qy      420 AlaIleLysLeu 423
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      44 GCGGTAAAGCTT 33

RESULT 4
AZ935488 544 bp DNA linear GSS 24-APR-2001
LOCUS      B2_Ba0003J24r B. japonicum BAC library Bradyrhizobium japonicum
DEFINITION genomic, genomic survey sequence.
ACCESSION  A2935488
VERSION     A2935488.1 GI:13778288
KEYWORDS   Bradyrhizobium japonicum
SOURCE      Bradyrhizobium japonicum
ORGANISM   Bradyrhizobium japonicum
REFERENCE   1 (bases 1 to 544)
AUTHORS     Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
            Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
            A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
            genome Res. 11 (8), 1434-1440 (2001)

TITLE       genome Res. 11 (8), 1434-1440 (2001)

JOURNAL     genome Res. 11 (8), 1434-1440 (2001)
MEDLINE     21376150
PUBMED      11483585
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: twing@clemson.edu
            Class: BAC ends
            High quality sequence stop: 499.

```

ORIGIN

```

/strain="USD110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/notes="Vector: pIndigo536; Site_1: HindIII"

```

Alignment Scores:

Pred. No.:	1,21e-42	Length:	708
Score:	514.00	Matches:	102
Percent Similarity:	58.05%	Conservative:	35
Best Local Similarity:	43.22%	Mismatches:	97
Query Match:	16.68%	Indels:	2
DB:	8	Gaps:	1

US-09-750-986D-30 (1-589) x A29334119 (1-708)

```

Oy 188 LeuPheThSerGlySerThrLyLeuProLyAlaValProThrThrGlnArgMetLeu 207
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 708 ATTTTCCTTAAGTCTNCGCGGGGCCCAAAAACCTTATCACACCCCAAGCATTTT 649
Oy 208 CysAlaAsnGlnImetLeuLeuGlnThrPheProValPhe-----GlyGluGluPro 225
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 648 GCCTTAGACACAGCCGCAAGGGGAGNCCCTTCTTAACAAGGGGTGGCAACAT 589
Oy 226 ProValLeuValAspTrpLeuProTrpPheAsnIleThrPheGlySerHisAsnIleGly 245
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 588 CTGCTCAATGTATACCTGGCTGCCCTTGACCCACACCTTGAGCGCAACCAATTTTCAC 529
Oy 246 IleValLeuTrpAsnGlyGlyThrTyTrpLeuAspAspLyLysProThrAlaGlnGly 265
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 528 CTGCTGTCGCCCAACGGCGCTCACTATATGACGGCGCAAGCGCTGGCGCGCTC 469
Oy 266 PheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyLeuThrValPro 285
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 468 TTTCGACAGCTGCTCGCGCAACCTGAAGAGCTGATGCGAGCTATTTCACATGTCGC 409
Oy 286 LyGlyTrpGluGluLeuValGlyAlaLeuGluLysArgSerThrLeuArgGluArgPhe 305
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 408 CGCGGCTTCGACATGCTGATCGCGCATTCGCGGATACCAAGCGTTGGCGCGCGCTTC 349
Oy 306 PheAlaArgMetLyLeuPhePhePheAlaAlaGlyLeuSerGlnGlyIleTrpAsp 325
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 348 TTTCAGCAGGTGAAGTTCGCTTCTTACGCGGCGCGAGCGCTCGCGCAACCTGTGGAC 289
Oy 326 ArgLeuAspArgValAlaGluGlnHisCysGlyGluArgIleArgMetLeuAlaGlyLeu 345
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 288 GCGCTCGAACAATGTCTCGTGAACCGTGGCGCGCGCTCGCATGTATGCGGCTCG 229
Oy 346 GlyMetThrGlnThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAlaGly 365
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 228 GGTTCGACGGAACCTTCGCGCGTGGCGGACCGCATTCCTCGCGGAGCGCTGGC 169
Oy 366 TyrIleGlyLeuProAlaProGlyCysGluValLyLeuValProValAspGlyLyLeu 385
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 168 AATATCGCGCTCGAATACCGGATACCGAGCTGAAGCTGTCACTCGCGCGCAACGCTG 149
Oy 386 GluGlyArgPheHisGlyProHisValMetSerGlyTyTrpArgAlaProGluGlnAsn 405
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 108 GAAGTGGCGCTGGCGCGCGCAACCTCACCGCGGTATACGGAAGCGCGCGAGCTGACC 49
Oy 406 AlaGlnAlaPheAspGluGluGlyTyTrpCysSerGlyAspAlaIle 421
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 48 AGCGAGCGGTTTGCACGAGGGCTTTATCTCATTCGGGAGTGGGCTG 1

```

RESULT 2
 A2933866/c 671 bp DNA linear GSS 24-APR-2001
 LOCUS B2_Ba0001M12r B. japonicum BAC library Bradyrhizobium japonicum
 DEFINITION genomic, genomic survey sequence.
 ACCESSION A2933866
 VERSION A2933866.1 GI:13775926
 KEYWORDS GSS.

SOURCE
 ORGANISM
 Bradyrhizobium japonicum
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE
 1 (bases 1 to 671)

AUTHORS
 Tomkins, J.P., Wood, T.C., Stacey, M.G., Job, J.T., Judd, A.,
 Golcochea, J.L., Stacey, G., Sadowsky, M.J. and Wang, R.A.
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum

TITLE
 genome Res. 11 (8), 1434-1440 (2001)

JOURNAL
 MEDLINE
 PUBMED
 11483585

COMMENT
 Contact: Ming RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Class: BAC ends
 High quality sequence start: 8
 High quality sequence stop: 625.
 Location/Qualifiers

FEATURES
 source
 1. 671
 /organism="Bradyrhizobium japonicum"
 /mol_type="genomic DNA"
 /strain="USD110"
 /db_xref="taxon:375"
 /db_host="E. coli"
 /clone_lib="B. japonicum BAC library"
 /note="Vector: pIndigo536; Site_1: HindIII"

ORIGIN

Alignment Scores:

Pred. No.:	1,86e-34	Length:	671
Score:	435.00 <td>Matches:</td> <td>94 </td>	Matches:	94
Percent Similarity:	57.92% <td>Conservative:</td> <td>34 </td>	Conservative:	34
Best Local Similarity:	42.53% <td>Mismatches:</td> <td>90 </td>	Mismatches:	90
Query Match:	14.11% <td>Indels:</td> <td>3 </td>	Indels:	3
DB:	8	Gaps:	1

US-09-750-986D-30 (1-589) x A2933866 (1-671)

```

Oy 203 ThrGlaArgMetLeuGlyAlaAsnGlnImetLeuGlnThrPheProValPheGly 222
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 663 ACCACGCGATGCTGACCTCGACGCGAGGCCAAGGCCAGACATTCCTTCGAA 604
Oy 223 Glu-----GluProProValLeuValAspTrpLeuProTrpAsnIleThrPheGly 240
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 603 CAGCGCGACGATCTGATCATTCGATGCTGCTGCGGACCGACCATTCGCGCGC 544
Oy 241 SerHisAsnIleGlyIleValLeuTrpAsnGlyGlyThrTyTrpLeuAspAspGly 260
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 543 AACCAACAATTTCAACTGCTGCGCGCAACCGGCGCTCACTATTCACACCGCGCAAG 484
Oy 261 ProThrAlaGlnIlePheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 483 CTGCGCGCGGCTTCNTGCGAGCTGCTCGCAACTGAAGCGTGATGACGAGCTC 424
Oy 281 TyrLeuThrValProLyGlyTrpGluGluValGlyAlaLeuGluLysArgSerThr 300
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 423 TATTTCAGATGCGCGCGGCTTCGACATGCTGATCAGCGCATTCGCGGTGACGAAGCG 364
Oy 301 LeuArgGluArgPhePheAlaArgMetLyLeuPhePheAlaAlaAlaGlyLeuSer 320
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 363 TTGCGCTCGCTTCTTCAGCGAGGAGTTCGCTTACGCGGCGCGCGAGCGCTGCGG 304
Oy 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGlyLysArg 340
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 303 CAGAACATGTGCGACCGCTCAACTATGTCTGCTGTAACCTGCGCGCGCTTGTCCCG 244
Oy 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPhe-ThrThrGlyPr 360
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 7, 2005, 23:57:11 / Search time 4035 seconds
(without alignments)
5319.207 Million cell updates/sec

Title: US-09-750-986D-30
Perfect score: 3082
Sequence: 1 MRSLEALPPGRIERLEH.....KVDALYRGEDQSMIRDEATL 589

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues
Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xidp
-O/cgnt 1/USPTO.spool_p/us09750986/runat_03022005_071744_18577/app_query.fasta_1.775
-DB=EST -QPMT=fastcap -SUFFIX=ret MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45
-DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09750986 @CGN 1.1 5180 @runat_03022005_071744_18577 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEJOURNEY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV_TIMEDOUT=120 -MARN_TIMEDOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: EST:
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gse1:*
9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	514	16.7	708	8	AZ934119 BJ_Ba000
C 2	435	14.1	671	8	AZ933866 BJ_Ba000
C 3	315	10.2	714	8	AZ933744 BJ_Ba000
4	301.5	9.8	544	8	AZ935488 BJ_Ba000
5	271	8.8	2698	3	AK054103 Mus muscu
6	264.5	8.6	2505	3	AK054103 Mus muscu
7	255	8.3	2848	3	AK004897 Mus muscu
8	253	8.2	1897	9	AY402446 Mus muscu
9	253	8.2	2548	3	CR606980 full-leng

10	250	8.1	2518	3	BC026161	BC026161 Homo sapi
11	244	7.9	2097	9	AY402444	AY402444 Homo sapi
12	244	7.9	2203	3	CR597120	CR597120 full-leng
13	232.5	7.5	2672	3	AK028078	AK028078 Mus muscu
14	226	7.3	2175	9	AY418305	AY418305 Homo sapi
15	226	7.3	2366	3	CNS0A7CK	EX824843 Arabidops
16	220	7.1	4243	3	AK034942	AK034942 Mus muscu
17	215	7.0	2166	9	AY418307	AY418307 Mus muscu
18	211.5	6.9	979	7	CK248477	CK248477 EST732114
19	201.5	6.5	1199	6	CD508894	CD508894 CDA93-D04
20	198.5	6.4	907	9	CG847891	CG847891 ZMMBBD031
21	197.5	6.4	2461	3	BC028399	BC028399 Homo sapi
22	197	6.4	829	6	CB675484	CB675484 OSUNE11E
23	197	6.4	831	6	CB675482	CB675482 OSUNE11E
24	194.5	6.3	909	7	CO114394	CO114394 GR_EB015
25	194.5	6.3	941	5	BX705748	BX705748 BX705748
26	190.5	6.2	898	7	CK260414	CK260414 EST706492
27	187.5	6.1	815	7	CO127557	CO127557 GR_EB12D
28	186.5	6.1	1034	8	B2676934	B2676934 PUBIF45TD
29	186.5	6.1	1265	6	CD503756	CD503756 CDA64-B06
30	185.5	6.1	2290	3	CNS0AARU	BX817150 Arabidops
31	185.5	6.0	909	7	CN157496	CN157496 946208 MA
32	184.5	6.0	766	6	CD445611	CD445611 EL01T0404
33	183	5.9	1811	3	CNS0AD10	EX818812 Arabidops
34	183	5.9	2977	3	AK012088	AK012088 Mus muscu
35	182	5.9	1833	3	CNS0A10J	BX832101 Arabidops
36	181	5.9	884	7	CK260415	CK260415 EST706493
37	180.5	5.9	810	7	CO200299	CO200299 GE02_6_D0
38	180.5	5.9	1841	3	CNS0ADKC	BX813547 Arabidops
39	180	5.8	1168	5	BX344418	BX344418 BX344418
40	179	5.8	931	7	CK245466	CK245466 EST729103
41	177.5	5.8	901	7	CK800668	CK800668 AGENCOURT
42	176.5	5.7	1920	3	CNS0A5CA	BX823446 Arabidops
43	176	5.7	796	4	BJ719420	BJ719420 BJ719420
44	175.5	5.7	1200	6	CD506373	CD506373 CDA78-H04
45	175	5.7	1832	3	CNS0AD12	BX814596 Arabidops

ALIGNMENTS

RESULT 1
AZ934119/c 708 bp DNA linear GSS 24-APR-2001
LOCUS BJ_Ba0001H18f B. japonicum BAC library Bradyrhizobium japonicum
DEFINITION genomic, genomic survey sequence.
ACCESSION AZ934119
VERSION AZ934119.1 GI:13776179
KEYWORDS GSS.
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE 1 (bases 1 to 708)

AUTHORS Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.

TITLE A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
Genome
JOURNAL Genome Res. 11 (8), 1434-1440 (2001)

JOURNAL MEDLINE 21376150
PUBMED 11483585

COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Class: BAC ends
High quality sequence stop: 661.
location/Qualifiers

FEATURES
source 1..708
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"

Db 361 ECEP--LVYGNPFDLIDYTAIGIPVSTREIVYDDAGNVVNDQVELVGRGQWQGY 418
QY 399 WRAPBONAQAFDEBGYYCSGDAIKLAD-----PADPOKGLMFD 436
Db 419 WQBPBATEKVELNAGMSTGDIYKFDDQGLIHIVRKKDMILVSGFNVYFNEIEDVALH 478
QY 437 GRAPBPKLSSGVFVSGPIRTAVLEGSGVVDVVVAAPDRBCLGLVFPRLDGR--- 493
Db 479 GKVLVNAALIQANDVSGELVYIKYVVRDPSLTVDVIAHCKRKHILTYKV--PKLVFPRDL 537
QY 494 ALSGLGK-----EASDAEVL 509
Db 538 PKTNVGIILRRVLRBEENDAQDLA 560

RESULT 15

E70937
probable fadD15 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C/Accession: E70937
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A/Authors: Squares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: E70937
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1600 <COL>
A/Cross-references: UNIPROT:O53521; GB:AL021957; GB:AL123456; NID:G3242293; PID:CAA1749
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: fadD15
C/Superfamily: Synchocytis long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homoid
F;66-585/domain: acetate-CoA ligase homology <ACL>

Query Match 8.4%; Score 257.5; DB 2; Length 600;

Best Local Similarity 21.7%; Pred. No. 3.3e-12;

Matches 132; Conservative 101; Mismatches 267; Indels 109; Gaps 22;

QY 34 RAANGERRISYEMFNVAIAQSLPYGLSARPLIYSGNDLEHQLAFGAMVAGIP 93
Db 37 RLIDGVTVDTCAEAAQIRAAALGLISGVQAGDRVIFSAITREYEMALIDFALAVG-- 94
QY 94 YCPVSPAYSLSDOLAKRIHIVGLQGLVFAADAPFORAIEITILPDVPAIFTRGELA 153
Db 95 -AVTVPTERTSSAE--QVRVLODSEAVVLPFAETDSHATWAE--LSGSVPALREVLOIA 149
QY 154 GRRTVSFDLSLEDPGIEADN-----APATGPDTIAKFLFTSGSTKLPRKAVFTQRL 207
Db 150 GSGGNALDRLTEAGASVDPALFTARLAALRSTDPATL---IYTSGTGRPKGCQLTQSNL 206
QY 208 CANOQMLQTFPPVGEPPVLVMDLPMNHFFGSGHNGIVLYNGCTYYLDDGKPTAQGFA 267
Db 207 VHEIKARAVHPTLLRGERLVLPLAHVL--ARALSMAAFH-----SKYTV--GFT 255
QY 268 ETLRN-----LSEISPTAYLVTKMBELVGLERDSTLRRFFARMKLPFAA----- 316
Db 256 SDIKLLPLMALAVKFTYVVSVPKVFKEKYNVTAEQNAANA---GKGRIFALMAQTAVDWS 311
QY 317 -----AGI-----SGIWMRL-----DRVABQHGERIRM 341
Db 312 EACDRGPGILLRAKHAVFDLIVRKLRALAGNCRAAVSGAPLGAFLGHFYRGAQLTI 371
QY 342 MAGIGMETAPSCFTTGPISMGYIGLPAKCEVXLVPVQKLEGRFHHGPHMWSGYRA 401
Db 372 YEGYISGTSIGVAISQFNDLKITGVKFPVGNLSRLAD--DGEILVR--GQVFSGTYRN 428
QY 402 PEONAAQFDEBGYYCSGDAIKLADPADPOKGLMFDGRIABFKLSSGVFVSGPIRTAV 461

Db 429 EQATTEAF--TDGWFXTGD-----LGAVDEBGFLLITGRKKELIYTAGG--KNVAPVLEBQ 481
QY 462 LEEGSYVLDDVVAAPDRBCLGLVFPRLDGRALSG-----LAKASDAEVL 509
Db 482 LRAHPILISQAVVVGDAKPFIGALI---TIDPEAFEGMKORNSKTAGASVGDLDATDPLLA 538
QY 510 SEVRAMFADWLKRLREATGNSRIIMWVGLDTPPSIDKGEVTDKGSINORAVLQWBSA 569
Db 539 EIDAA-----VKQANL-AVSHAESIRKPRILFVDFTEDTGELTPMKVARKVAEKFA 591
QY 570 KYDALYRGE 578
Db 592 DIEAIYNKE 600

Search completed: February 7, 2005, 23:50:01
Job time : 42 secs

Db 148 SIRVOLLANDSGEHNLODAINHPADGFTAPPSPADDEVVYFQLSGTTGTPKLIPTRTNDY 207
Qy 205 -----RMLCANQOMLLQTFPVGGEPPVLDVLPNNHTFGSGH--NIG 245
Db 208 YSVRSRVEICQFTQCRYLCA-----IPXHNHYANSSPGSLG 245
Qy 246 IVLNGSTYVLDGKPTAQAFAETL-----RNLSISPTAY-----LTVPKMEE 290
Db 246 VFL-AGGTAVL--AADPSA-----TLCFPLIEKHQINTALVPAPVSLMLQALTEGESRAQ 298
Qy 291 L-----VGLARDSTLRERFFARMKLFPPAAAGLSOGI--WDRLDRAVNEOHCGERIRM 341
Db 299 LASLKLQVGAARLSATLAARIPAEICQLOQVFGMAEGLVNTYRLDLSAEK----- 350
Qy 342 MAGLGMTERAPSCFTFTTGPLSMAGYIGLPAPGCEVKLVDPDGKLEGFHGFPHWMSGYRA 401
Db 351 ---IHTQGYPMC--PDDEVVADAEENPLPQGEV-----GLMTR--GPYTFRGYKKS 397
Qy 402 PEONAOAFDEEGYCGSDAIKLADPADPQKGLMFDGRIADFDKLSGCVFVSGPLRTTAV 461
Db 398 POHNASAFDANGFYCSGDLISI-----DREGYITVQGREKQ-----I 435
Qy 462 LEGGSYVLDVVAAAPDEECGLVFPRLDCRALS---GLKESADAEVLASEPVRAMFA 518
Db 436 NRGE-----KIAAEETENL-LRHPRVITAAVLSMEDELMEGKSCAYLVKPELR---A 486
Qy 519 DMLKRLNREATGNASRIMVGLDTPPSIDKEVTGKSINORAVLQWRSAYDA 573
Db 487 VQVRRLFRE-----QGIAEFKLPDRVCEVDSLPLTAVGVKQKQLRWMLASRASA 536

RESULT 13
A99708
2,3-dihydroxybenzoate-AMP ligase [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: A99708
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
R:gaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Ref. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: A99708
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-536 <NAV>
A:Cross-references: UNIPROT:Q8XBVJ; GB:BA000007; PIDN:BA034056.1; PID:913360091; GSFDDB:
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECG0633
C:Superfamily: 4-coumarate-CoA ligase; acetate-CoA ligase homology

Query Match 8.4%; Score 260; DB 2; Length 536;
Best Local Similarity 24.0%; Pred. No. 1,8e-12;
Matches 143; Conservative 77; Mismatches 215; Indels 160; Gaps 27;
Qy 37 NGEWRISYAEPMHNVAIAQSLIPYGLSAERPLLIVSGNDLEHLLQAFGAMTAGIYCP 96
Db 44 DGB-RQLSTYRELNOADNLACSLRQCIKRGERTALVGLNAVLEYITFFALKLG----- 97
Qy 97 VSPAYSLISODLAKLRHIVGLQGLVFAADAAPFORAIEITLPDDVPAIFRTGELAGRR 156
Db 98 VAVVLLALFHSQRSELNVAASQIEPALIIA-----DRQHALFSGDD---FLMTFVAEHS 147
Qy 157 TVSFDSLLEPG-----GIEADNAFAATGP-DTIAKLFRTSGSKLPRKAVPTTQ--- 204
Db 148 SIRVOLLANDSGEHNLODAINHPADGFTAPPSPADDEVVYFQLSGTTGTPKLIPTRTNDY 207
Qy 205 -----RMLCANQOMLLQTFPVGGEPPVLDVLPNNHTFGSGH--NIG 245
Db 208 YSVRSRVEICQFTQCRYLCA-----IPXHNHYANSSPGSLG 245
Qy 246 IVLNGSTYVLDGKPTAQAFAETL-----RNLSISPTAY-----LTVPKMEE 290

Db 246 VFL-AGGTAVL--AADPSA-----TLCFPLIEKHQINTALVPAPVSLMLQALTEGESRAQ 298
Qy 291 L-----VGLARDSTLRERFFARMKLFPPAAAGLSOGI--WDRLDRAVNEOHCGERIRM 341
Db 299 LASLKLQVGAARLSATLAARIPAEICQLOQVFGMAEGLVNTYRLDLSAEK----- 350
Qy 342 MAGLGMTERAPSCFTFTTGPLSMAGYIGLPAPGCEVKLVDPDGKLEGFHGFPHWMSGYRA 401
Db 351 ---IHTQGYPMC--PDDEVVADAEENPLPQGEV-----GLMTR--GPYTFRGYKKS 397
Qy 402 PEONAOAFDEEGYCGSDAIKLADPADPQKGLMFDGRIADFDKLSGCVFVSGPLRTTAV 461
Db 398 POHNASAFDANGFYCSGDLISI-----DREGYITVQGREKQ-----I 435
Qy 462 LEGGSYVLDVVAAAPDEECGLVFPRLDCRALS---GLKESADAEVLASEPVRAMFA 518
Db 436 NRGE-----KIAAEETENL-LRHPRVITAAVLSMEDELMEGKSCAYLVKPELR---A 486
Qy 519 DMLKRLNREATGNASRIMVGLDTPPSIDKEVTGKSINORAVLQWRSAYDA 573
Db 487 VQVRRLFRE-----QGIAEFKLPDRVCEVDSLPLTAVGVKQKQLRWMLASRASA 536

RESULT 14
H82132
long-chain-fatty-acid-CoA ligase VC1985 [imported] - Vibrio cholerae (strain N16961 serog
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: H82132
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gynn, M.L.; Dodson, R.J.;
R:chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellers, P
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Frazer, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: H82132
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-566 <HEI>
A:Cross-references: UNIPROT:Q9KQJ3; GB:AE004273; GB:AE00852; NID:9656517; PIDN:AAE9513;
A:Experimental source: serogroup O1, strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1985
A:Map position: 1
C:Superfamily: 4-coumarate-CoA ligase; acetate-CoA ligase homology
P:70-550/Domain: acetate-CoA ligase homology <ACL>

Query Match 8.4%; Score 258; DB 2; Length 566;
Best Local Similarity 24.3%; Pred. No. 2,7e-12;
Matches 122; Conservative 69; Mismatches 218; Indels 94; Gaps 18;
Qy 76 NDEHQLAFGMYAGIIPYCPVSPAS-----LSQDLAKLRHIVGL 118
Db 83 NLLQYVALFGVRAQMIANVAPLTPPRELHQLNDADARAIVISNPANTLEQIVANT 142
Qy 119 Q-EGVFAADAAFORAIEITLPDDVPAIFTRG-----ELAGRTVSFDLSLEPGGIEA 172
Db 143 QVGHVVLTSIGMLPPAKGIV--DVVKYKGMVRYKIDLPG--AISMRAALKHGRRLQY 198
Qy 173 DNAFAATGPDITIAKLFRTSGSKLPRVPTTQRMLCANQOMLLQTF-----PVGGEPPVL 228
Db 199 VKPF-MSGED-IAFLQYTGTTGVAAGAILTHRNVAAN---VLQAGAYGFPVLOEGRELV 253
Qy 229 VDMLPNNHTFGSGSHNIGIVLVNGCTYYLDDGKPTAQAFAETLRNLSEISPTAYLVTPKGM 288
Db 254 VTALPYHVFALVNCLEFLEMGRMLITNPRDIDGFVKELOKYP-----FTAITGV 306
Qy 289 EELVGLERDSTLRERFFARMKLFPPAAAGLSOGIMDRLDRAVNEOHCGERIRRMAGLGT 348
Db 307 NTLFNLVNNEDHELDPRMKLAVGGMVQRAVVERMKKTGVH-----LLEGYGLT 360
Qy 349 ETAPSCFTTGPLSMAGY---IGLPAPGCEVKL-----VPVDGKLEGFHGFPHWMSGY 398

```

Db 202 -----ANVIVDT---OKOLEKILIKWDLPHLKAUVIYOEPKKNANYTMBELI 250
Qy 154 GRRTVFSDLLEPGIEADNAPAAAT-GPRTIAKFLFTSGSTKLPKAVPTTQRMLC----- 208
Db 251 E-----LGOVPEBALDAILDIOQFNOCVLYVTSGTGNPKVMLSODNIWTAR 301
Qy 209 -ANQOMLLQTFPFGEPPVLDMLPMNHTFGSGHNIGIYLVNGTYYLDDGKPTAAGFA 267
Db 302 YGQAGDIOQAEV---QOEYVVSYLPLSHIAAQIYDLMGTGQMAOUCFPDPAKKTIV 358
Qy 268 ETLRNLSISPTAYLVTPKGMELVGLERDSTLERFPFAMKLPFAAGLSGCI----- 323
Db 359 NTLR---EVEPTSHMGVPRVWEKIMERIO-EVAASGFIRKMLMAMSVTLLEQNLCPSS 414
Qy 324 -----MDRLDRVAEQ-----HCGEILRMAGIGMT 348
Db 415 NDKPFTSLADYLVLARVROALGPAKCKQKNFYGAADMTAETORFPLGLNIRLTAAGLS 474
Qy 349 E-TAPSCFTTGPLSMAGY-IGLPAPGCEVVLV--PVDGKLEGRFHGPHVMGWRAPAQ 404
Db 475 ESTGPH--FMSSPYNVLYSSGRVPCRYKLVNQDADGICLMGRTTFMGILNMDK 532
Qy 405 NAOAFDEEGYCGSDAIKADPADPQKLMFDGRIADFKLSSGVFVSGPLRTAVLEG 464
Db 533 TCERIDSEGMVHTGDMQRL-----DADGFLYITGRKELIITAGGENVPVPIEEAVGME- 587
Qy 465 GSIVLDVVVA-----PRECLG--LVFPRLLD-----CRALSGLOKEA 502
Db 588 ----LPIISSAMLIGDQKFLSMILTKCTLDPEPTSDSLTEQAVEFCORVSKASTV 643
Qy 503 SDAEVLASEPVAWAFDMLKRLNREATGNASRIWVGLDTPPSIDKEVTDKGSINQRA 562
Db 644 SEIVGQDEADVQAIHIGICRVANANNAAPRYHIQKMAIIGDRPSISGSEGLPMKLRIT 703
Qy 563 VLGMSAKVDALYRGEDQ 580
Db 704 VLRRYKDIIIDSFYOEQKQ 721

```

RESULT 11

```

H69274
Probable long-chain-fatty-acid-CoA ligase (EC 6.2.1.3) AF0200 [similarity] - Archaeoglob
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: H69274
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kittness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: H69274
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-601 <KLE>
A:Cross-references: UNIPROT:O30039; GB:AE001092; GB:AE000782; NID:G2689415; PIDD:AA9103
C:Superfamily: Synecocystis long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homolo
C:Keywords: acid-thiol ligase; coenzyme A
F:65-587/Domain: acetate-CoA ligase homology <ACL>

```

Query Match 8.6%; Score 265; DB 2; Length 601;

Best Local Similarity 21.4%; Pred. No. 8.4e-11; Indels 122; Gaps 21;

Matches 131; Conservative 106; Mismatches 253; Indels 122; Gaps 21;

```

Qy 37 NGEWRISYAEFHNRAIAQSLLPYGLSARPLLIVSGNDLEHLQALFGAMVAGIYCP 96
Db 39 DGRKLVLYTEFMRVWRVRLSKFLISSGLRKGDRAIYADRWEIADPAVLTAG---GV 95
Qy 97 VSPAYSLSDLAQLRIHIVGLQPGVLPADAAFPQAIETILPDDVPAIFTRGELAGR 156
Db 96 VLVTVHSLNRE--QVEYILRDSERVVFT-----EKYAVNVEDEDEVFLR----- 140

```

```

Qy 157 TVSFDSLEPGIEADNAPAAATGPRTIAKFLFTSGSTKLPKAVPTTQRMLCANQOMLLQ 216
Db 141 --ELEKLVGEVSDDEESRKSVEPDDLASIVTTSSTGEPKAMLTMMWRNCHSVNS 198
Qy 217 TFPVFEPPVLDMLPMNHTFGSGHNIGIYLVNGTYYLDDGKPTAAGFAETLRNLSEI 276
Db 199 ITFYPGEPIHC--YLPLSHVQ-----RLVFPAGISRAATAVFCSPQGLFT---STAV 248
Qy 277 SPTAYLTVPKGME---LVGALERDSTLERFPFAMKLPFAA-----AAGLSQ 321
Db 249 KPVGLVPRILERRVAVGIVEKEKSPALAK-----KIFYSRGVAIEGKMSRGEKX 302
Qy 322 GIV-----DRL-----DRVAQCHGERIRRMAGIGMT 348
Db 303 GFWLNKRIIADLVSKIRENLGLTRIRFVCSAAELQELAYMNGCIPVIEGIGMT 362
Qy 349 ETAPSCFTTGPLSMAGYIGLPAPGCEVVLVVDGKLEGRFHGPHVMGWRAPAQ 408
Db 363 ETAPSNLNVGRFPGQVGPPIGIE-EAIAEDGELVR--GDNVKKGWRKEAETRTKT 419
Qy 409 FDEEGYCGSDAIKADPADPQKLMFDGRIADFKLSSGVFVSGPLRTAVLEG 468
Db 420 FTEDGMLKTGDLGEF---DEDGVLVFLGRKKHIVLDTGKNVSPVIEEB--LKNPLV 473
Qy 469 LDVVVAAPDECTGLVFPF--LLDORALSGL-----GKE---ASDAEVL 508
Db 474 SDAVIIGDGKPYTTILVPRFSLFADKNGIEYRSRTIYVRSISGEELIYAVDENLV 533
Qy 509 ASEPVRAWFA---DWLKLRLNREATGNASRIWVGLDTPPSIDKEVTDKGSINQRAVL 564
Db 534 ENNAVKELYAKIVDVNSRLAKHET-----IKFKILPEAFSLKEKEITPPLKRRHVL 588
Qy 565 QMSAKVDALYR 576
Db 589 KRYEKPIEENYK 600

```

RESULT 12

```

E85558
2,3-dihydroxybenzoate-AMP ligase (imported) - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 01-Mar-2002
C:Accession: E85558
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
.; Miller, L.; Grobbeck, B.O.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85558
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-536 <STO>
A:Cross-references: GB:AE005174; NID:G12513487; PIDD:AA654929.1; GSPDB:GN00145; UWGP:207;
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: entB
C:Superfamily: 4-coumarate-CoA ligase; acetate-CoA ligase homology

```

Query Match 8.5%; Score 261; DB 2; Length 536;

Best Local Similarity 24.0%; Pred. No. 1.5e-12;

Matches 143; Conservative 77; Mismatches 215; Indels 160; Gaps 27;

```

Qy 37 NGEWRISYAEFHNRAIAQSLLPYGLSARPLLIVSGNDLEHLQALFGAMVAGIYCP 96
Db 44 DGE-RQSLSTRELNQADNLCSLRQGIKRGETALVQGLVAVLYTTPALLKLG----- 97
Qy 97 VSPAYSLSDLAQLRIHIVGLQPGVLPADAAFPQAIETILPDDVPAIFTRGELAGR 156
Db 98 VAVVLAFLFSHQRELVNAYVQSIEPALLIA-----DRQALPFGDD-----PLNTFVAHS 147
Qy 157 TVSFDSLEPG-----GLEADNAPAAATGP-DTIAKFLFTSGSTKLPKAVPTTQ--- 204

```

F;56-583/Domain: acetate-CoA Ligase homology <ACL>

Query Match	9.2%;	Score 283.5;	DB 2;	Length 601;
Best Local Similarity	21.8%;	Pred. No. 2.9e-14;		
Matches 136;	Conservative 94;	Mismatches 252;	Indels 143;	Gaps 24;

```

QY 37 NGEMRRISYAMFHNVRALIAOSLIPYGSARERPLIYSGNLEHQLQAFGMVAGIPYCP 96
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
Db 30 NGOMOALIMHOPGEOLDELSWALLAOGIGVODKTAIFSNMPPRTIADFAALQIRAVTVP 89
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
QY 97 VSP-----AYSLSODLAKLRHIVGLLOGLVFAADAAEFORAIETILBDVPAPAFTR 149
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
Db 90 IYPTNTEPQAAYIIIONADV-----VVFVGQAQFDALISGF--EQCPBL--R 133
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
QY 150 GELGRRRV-----FDSLEOPGGIEADNAFAATGPDTIA 185
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
Db 134 LIVAMNANIDIKOASCAMHWDEFVAOSHODRAPLDALIEO-----ANYDLF 181
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
QY 186 KFLFSTGSTRPKAVPTTORMICANQOMLQTFPVGGEPPVLVDWLPAMHTFGSHNG 245
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
Db 182 TLIYTSSTGTGPKGVMIDYRNIGQLEBGHODRLNLTQED--VSLCFPLSHVERAWT--A 238
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
QY 246 IYLYNGGT--YLLDDGKPTAOGFAETLRNLSEISPTAYLTVEKGMEELVGALESDTIRER 304
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
Db 239 YLYKKGATNICYLO-----VAHVREALAEVRPTVMCAVPRFEKIPSAIH--EKVKA 289
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
QY 305 PFAPMKLFFFPAA--AGLSGSI-----W-----DPL-----DVAEHC 335
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
Db 290 PLVKKVLTFTAMVNNGAKMAVCRQOORPSMKKOSHOPADLTLYLSKLRALIGGINEMPC 349
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
QY 336 GER-----IRMMAGLMTETABSCFTTGPGLSMAGYIGLPARGCVKLVPV 381
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
Db 350 GGALDLETIGFFHAIGINVTLGYGMITETTATVSCWDHCHNPISIGISMGAQVK---I 406
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
QY 382 DGKLEGRFHGNVSGTIRARPEQNAQAFDEBGYICSGDAITLADPADPQKLMFDGRIAE 441
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
Db 407 GENNEIIVRGSMVWVRGYKLDKETTESFDEBGFLTKTGAGHI---DENGMLFTDRIKE 462
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
QY 442 DFKLSGGEVFEVGLRTRAVLEGG---SYULDVVVAAPDESCGLLVFPRLLDC--RALS 496
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
Db 463 LMKTSGGKYIA-----POVIEGAIGKDHFEQIAVLNDRTKFVSALVP---CPDSLE 512
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
QY 497 GLGKEAS-----DAVILASEPVRAMFADWLRLNREATGNASRIIMVGLDTPPSIDKE 551
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
Db 513 EYAKELNIKYHDRIELIKHSQVLEMFEXKRVNELQKE--LAKEQYKKRPLRPKARSMSGE 571
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
QY 552 VTDKGSINQRAVLQWRSAKVDALYR 576
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
Db 572 LTPQKLRKRVINDRYODEIEBMYQ 596
   | | | : : : : : | | | : : : | | | : : : | | | : : : | | | : : :

```

RESULT 9

T07944

C/Species: Brassica napus (rape)
probable long-chain-ally-acid-CoA ligase (EC 6.2.1.3) - rape

C/Date: 21-May-1999 #sequence_revison 21-May-
C/Accession: T07944
R:Fulda, M.

submitted to the EMBL Data Library, April 1996
A:Reference number: Z16223
A:Accession: T07944
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA
A:Residues: 1-677 <FUL>
A:Cross-references: UNIPROT:O96338; EMBL:Z72152; PDB:1CA95622.1
A:Experimental source: cv. Ascarai; developmental stage: 20-25 days after flowering; tissue: leaf
C:Superfamily: human long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homology
C:Keywords: acid-thiol ligase; coenzyme A
C:135-670/Domain: acetate-CoA ligase homology <ACL>

Query Match 8.9%; Score 273; DB 2; Length 677;
Best Local Similarity 24.5%; Pred. No. 2.3e-13;
Matches 121; Conservative 74; Mismatches 175; Indels 124; Gaps 24;

```

QY 38 GEMRRIISVYEMHNHRAIKQSLIPGLSABRPLLIYSGNLEHU---QLARGAMVAGI.PY 94
Db 110 GYKMMNTTGEVGTARTALSSGLVHHGITTGSSVIGIYINPEMLIYDQACSYSYSV-- 167
QY 95 CEVSPAYSLLSODLAK-LRH-----IVGLIQGLVFADAAPFORAIEITI--LPDDV 143
Db 168 ----FLYDTLPGDVAVFIVNHANVQAIFCVAETIINLSLGLSEMPSVRLVVVVGILNESL 223
QY 144 PAIFIRGELAGRRYTSFEDLLBPGGLEADNAPATGPDTIILAKLFTSGSKLPRAVETT 203
Db 224 PSL---PSSAGKVVVSYSVLLNO--GRSNPQPCPPKPDVAITLYTSGITGTIPGAVLT 278
QY 204 QRLCANQOMLQTEPEVFESEPPVLVDMIPMNHTFGGSHNIGIYLVNG--GTYIYLDGK 260
Db 279 HANLLIAN--VAGSSFSVKFFSSDIYISYLP LAHYERANQI-LAVYFGVAVGFQGDVWK 335
QY 261 PTAQGAETLRNLSEISPTAYLTVP---KGMELVALERDSTLRERFF----ARKKL 311
Db 336 ----LDDLLAALRPITVFSVPRLYNNIYIGITNAVVTSGGLKERLFFNAAYNAKQA 387
QY 312 PFPAAAGLSQGIWMDRL-----DRVAEOHCGERIRPMA----- 343
Db 368 LLNGKS--ASPIIDRLRVFNKIKIDRL-----GGRVRFPTSGASPLSPEVLEFLKICFGGRV 440
QY 344 --GLQMTETAPSCFTF---TGPLSMAGYIGLPAPGCEVKTLPVD-----GKL 385
Db 441 SEGVMETTER--SCVISGMDEGD-NLTGHVGSPPAPACEVTKLMDVPEMNYTSADDEPHRGEI 497
QY 386 EGRFHGPHMSSGYMRAPEQNAOAFDEEGYCSGDPAIKLADPADQKGMIMFGD-----R 438
Db 498 CYR--GPITFRGYKDEVQTRBVIDEDGMLHTGDI-----GLMLPGRLKIIDR 544
QY 439 IAEDFKLSGVEVFS 452
Db 545 KKNIFKLAGGEYIA 558

```

RESULT 10

JC7557

lipidosin - mouse
lipidosin-related protein
N;Alternate names: lipidosin-related protein

C1:Species: Mus musculus (house mouse)
C1:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 09-Jul-2004
C1:Accession: J07557
R1:Morita-Saeo, A.; Hida, A.; Inagawa-Ogoshiwa, M.; Waga, M.R.; Sugiyama, K.; Shimizu, J.
Biochem. Biophys. Res. Commun. 279, 62-68, 2000
A1>Title: Novel acyl-CoA synthetase in adrenoleukodystrophy target tissues.
A1:Reference number: J07557; UID: 20563802; PMID:11112418

A:Contents: Brain
A:Accession: JG7557
A:Molecule type: mRNA
A:Residues: 1-721 <MOR>
A:Cross-references: UNIPROT:Q99PU5
C:Comment: This protein possesses long-chain acyl-CoA synthetase activity and catalyzes the impairment of fatty acid metabolism in X-linked adrenoleukodystrophy. This in human X-linked adrenoleukodystrophy.
C:Keywords: adrenoleukodystrophy

Query Match	8.8%	Score 270;	DB 2;	Length 721;
Best Local Similarly	22.7%	Pred. No. 4,4e-13;		
Matches 154; Conservative	87;	Mismatches 287;	Indels 150;	Gaps 24

[illegible]

C:Genetics:

A:Gene: VP00537

C:Superfamily: Synecocystis long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homolo

Query Match 9.4%; Score 289; DB 2; Length 601;

Best Local Similarity 21.3%; Pred. No. 1,1e-14;

Matches 135; Conservative 115; Mismatches 268; Indels 116; Gaps 22;

```
QY 14 ILERLHMAKTRPQTCVAAARAANGEMRRISYAEFNNVPAIAQSLLPYGLSARPLLIV 73
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 10 LVRLRQOVSDRSELVAFREMSPEGE-KOLNWOODIDHVTIRISTALLSLAAIOERIGIF 68
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 74 SGR-----DLHLQAFGAMVAGIPYCVSPAYSLLSODLAKRIHIVGLQGLVFAA 126
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 69 ANNSMMSLVDLAILDL-----RGVSV-----PLIYA--TNTAAQAYIVADAVRIILFVG 116
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 127 DAAPFORAI--ETILPDDV-----PAIFTRGELAGRTVSPDSLEBQPG----- 168
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 117 EGQOFDAITLQPLCPQLAHIIIVLDPVADLRGCEYAGHLADFEG-QQQPDTVOQHLLTTR 175
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 169 --GLEADNAPAAAGPDTIAKFLFTSSGSKLPKAVPTTQRMLCANQOMLLQTFPVGSEEP 226
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 176 IECDCDDDLFT-----LIYTSGTGEPKGMLDYRMMAAQLYHLDRLTLTPGD-- 224
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 227 VLVDMLPMNHTFGSGSHNIGIVLYNGTFYDDGKPTAQGFAETLRN-LSEISPTAYLTVP 285
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 225 VSLSFLPLSHVFERAMSFYVMHTGAQNVYTRN-----TDVYRSAMQAVKPTVMCAVP 276
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 286 KGMEEVLGALERDSTLEREPFARMKLFPPAAG-----LSQGIWDRLDYV 330
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 277 RPYEKVPSAI-NDKVALAKMHQRYLPHMAVGCGRKRPLOQGRPLPWLSEQMYTLADRL 335
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 331 AEQH-----CGERIRMA-----GLGTEPARPCTFTTGLSM 363
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 336 VLKRLGVLGGRVRFPLPAGARLDNIIILFQALGVNIKKYGYGTEFCATVSCWEEQNR 395
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 364 AGYIGLPAPGCEVVLVVDKLEGRFHPHWSGYWPAPEONAOAFDEGYVCSGDAIKL 423
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 396 FGSLGKRLPGIDVRL--GAENEIQAQVGYVMKGYFKKPOETLEAFEDGMVKTGIDGAL 452
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 424 ADPAPOKGLMFDGRIAEDFKLSSGVFSVGPRLTRAIVLGGSYVLDVVVAADRECLGL 483
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 453 ----DAQHILFITERLKDLMKTSQGYIA--PQWIEGTGADPRIEQVALIADTRKVSVA 506
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 484 LVPRRL--LDCRALSGIGKASDAEVLASEPVAMFPMKRLNREKTAGASRIWVGL 541
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 507 LIVPSESLEEVARSINLKYHDLRLHSHIVMFEORLKEIOKE-LALFEQVKKFTLL 565
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 542 DTPPSIDKGEVTDKGSINQRAVLQWRSKAYDALY 575
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 566 PQAFTMETGELTPMKLRKIIIDORYONEIDLMY 599
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 7

H64041

probable long-chain-fatty-acid-CoA ligase (BC 6.2.1.3) H10002 [similarity] - Haemophilus

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004

C:Accession: H64041

R:Heidechmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Pine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: H64041

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-607 <TIGR>

A:Cross-references: UNIPROT:P44446; GB:U12686; GB:I42023; NID:g3212175; PIDN:AAC21681.1;

C:Superfamily: Synecocystis long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homolo

C:Keywords: acid-thiol ligase; coenzyme A

F:63-590/Domain: acetate-CoA ligase homology <ACL>

Query Match 9.3%; Score 286; DB 2; Length 607;

Best Local Similarity 21.0%; Pred. No. 1,9e-14;

Matches 137; Conservative 115; Mismatches 251; Indels 150; Gaps 28;

```
QY 15 ILERLHMAKTRPQTCVAAARAANGEMRRISYAEFNNVPAIAQSLLPYGLSARPLLIVS 74
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 16 VRIHQOAKTRMTMTLYRK-ENGLMRDISMKRFGQQLNLSALAHNIIDVQDKAIFR 74
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 75 GN-----DLHLQAFGAMVAGIPYCVSPAYSLLSODLAKRIHIVGLQGLVFAAD 127
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 75 HNNERTWITDIATLQIR-----AITVPYATNTAQAEF--ILNHADVKILFVGD 122
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 128 AAFQRAIFETILPDDVPAIFTRGELAGRTVSPDSLEBQCGLEAD-----NAFATGP 181
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 123 QEOYDTLE-IANHCPRK-----QKIYAKSTIQ---LQDPLSCWESFKTQS 168
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 182 DT-----IAKFLTSGSTRKLPKAVPTTQRMLCANQOMLLQTFPVGSEEP 225
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 169 NAQODELTQRLNQKSLDLETTIYTSGTGEPKGMWLDYANLA--HQLETHDLSLWTDQ 226
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 226 PVLVDMLPMNHTFGSGSHNIGIVLYNGTY-YLDDGKPTAQGFAETLRNLSISPTAYLTV 284
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 227 DISLSPFSHIFERAM-AAVILHRCALCYLEDTNQVBSA-----LTEIRPTLMCAV 278
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 285 PKMEELVGA---LRDSTLERFPARMKLFPPAAGLSQGIWD-RIDRVA----- 331
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 279 PRYERKIYAAVLDKVKAPRLRQ-----IMFWAISVGQYFDLRANKAIPLLIKQ 331
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 332 -----BOHCERIRMA-----AGIMETAB--SC 354
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 332 PALADKLVSLSKRLQLLGRIKQWPCGALERPALIGFPHAIGINILGGMETTATVGC 391
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 355 --TFTTGPLSMAGYGLPAPGCEVVLVVDKLEGRFHPHWSGYWPAPEONAOAFDEE 412
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 392 WHDFQNPNS-----IGTLMPKAEVK--IGENNEILVRCGMWKKGYKKPEETQAFTED 444
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 413 GYVCSGSDAIKLDPAPOKGLMFDGRIAEDFKLSSGVFSVGPRLTRAIVLGGSYVLDVY 472
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 445 GFLKTDABEF--DEQNLFTTDRIKELMKTSNGKYIA--POYTESKIGDKXFEQIA 498
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 473 VAAPDECEGLIVFPLRLDC-RALSGIGKEAS-----DAVLASEPVAMFPMKRLNR 526
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 499 IYDACKVYSALIV---CPDSLEEVAKQMINIKYHDLRLKNSIILMGFERIAIVAQK 554
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 527 EATGNASRIWVGLDTPPSIDKGEVTDKGSINQRAVLQWRSKAYDALYGED 579
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 555 E-LAHFEQVKKFTLLSQAFSIKLGEITPTLKLRRKVIILRRYRQIEMVHNSQE 606
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 8

H82069

probable long-chain-fatty-acid-CoA ligase VC2484 (imported) - Vibrio cholerae (strain N1

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004

C:Accession: H82069

R:Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chaudson, D.; Ermolaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellers, P. Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: AB2035; MUID:20406833; PMID:10952301

A:Accession: H82069

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-601 <HEI>

A:Cross-references: UNIPROT:Q9KP89; GB:AE004318; GB:AE003852; NID:g9657063; PIDN:AAF9562

A:Experimental source: serogroup O1, strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2484

A:Map position: 1

C:Superfamily: Synecocystis long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homolo

RESULT 4

H87662

fenuoyl-CoA synthetase [imported] - *Caulobacter crescentus*C/Species: *Caulobacter crescentus*

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

C/Accession: H87662

R/Merman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eissen, J.; Heidelberg, J.

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kjol

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of *Caulobacter crescentus*.

A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: H87662

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-621 <STO>

A/Cross-references: UNIPROT:Q9A368; GB:A8005673; NID:gl3425038; PIDN:AAK23300.1; GSPDB:C

C/Genetics:

A/Gene: CC3338

Query Match 31.5%; Score 972; DB 2; Length 621;
Best Local Similarity 38.3%; Pred. No. 1.5e-68;
Matches 219; Conservative 86; Mismatches 249; Indels 18; Gaps 6;

```
QY 11 PGRILRLERHMAKTRPQTCVAAARAAN-GEWRRI SYAEMFNVAIAQSLPLPGLSAERP 69
DB 50 PRSIAHLIAAKAEHPDRPYLKQREPHNGPMRGVTEGARHAYBGVAQWMLDKLQPHDS 109
QY 70 LLIVSGNDLHLQLAFAMTAAGIPYCPVSPAYSILSODLAKLRIHIVGLQGLVFAADAA 129
DB 110 VMLISGSIHAAWMTLCAAYAGVPAAPISPAVSLISTDHAKLKCFEKVAPRVVFAQSGA 169
QY 130 PQRARITLIDDDVPAIFTEGELAGRTVSFDSLLEQPGIEADNAAAGSPDTIAFLF 189
DB 170 MFGALATLIALDPSLVITTAADGAGGATTFEAATAATTPTAAVAADAGPATVAKTLF 229
QY 190 TSGSTGLPKAVPTTORMLC--ANQWMLQTFPVFGESEPVLYDMLPMNHTFGSHNIGI 246
DB 230 TSGSTGLPKAVPTTORMLC--ANQWMLQTFPVFGESEPVLYDMLPMNHTFGSHNIGI 287
QY 247 VLYNGGTYVDDGKPTAOGAETLRNISEISPTAYLTVPKGMBELVGLERDSTLRERFF 306
DB 288 VINGGTYVDDGKPTAOGAETLRNISEISPTAYLTVPKGMBELVGLERDSTLRERFF 347
QY 307 ARKKLFFPAAGISQGIWDLDRVAEQHGERIRIMMAGLMTETAPRSCFTTGTGLSAGY 366
DB 348 KMLRYNGYGAATISNDYERIKQALAVAFTHRIPLITMYGATET-QGIIIVHWITERVGL 406
QY 367 IGLPAPGCEVAVLVVDGKLEGRFHGHVMSGYWRAPQNAQAPEDEGGYSCGDAIKLADP 426
DB 407 VGLPLPGLQKLAAPSQSGKYEVAVGVVAAAGYHNDPRTAAAPDEEFYGLGDAARVDP 466
QY 427 ADPQKGLMFGRIAEDEKLSGGVFSVGLPRTAIVLEGGSYVLDVVAAPDEEGLGLV 486
DB 467 DDPAGKGLVFGRIAEDEKLSGGVFSVGLPRTAIVLEGGSYVLDVVAAPDEEGLGLV 526
QY 487 PRLLDCAALSGLGKAEADAEVLASEPV---AMPADALXRLNREAGTANASRIMVGLDIT 543
DB 527 P-----SPAGLAALVADPGECTPLEKVAIIKQDLAFLNAAAGSSRRVAFETILTE 578
QY 544 PSIDKGEVTDGKSGINORAVLQWRSKADALY 575
DB 579 PSIDAGEITDKGIVNRCVLERADRVVALLY 610
```

RESULT 5

E87040

acyl-CoA synthase [imported] - *Mycobacterium leprae*C/Species: *Mycobacterium leprae*

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

C/Accession: E87040

R/Cole, S.T.; Elgimeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho

R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Felwell, T.; Fraser, A.; Hamlin, N.; Holroyd,

eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A/Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squ

A/Title: Massive gene decay in the leprosy bacillus.

A/Reference number: A86909; MUID:21128732; PMID:11234002

A/Accession: E87040

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-476 <STO>

A/Cross-references: UNIPROT:Q50017; GB:AL450380; NID:gl3093059; PIDN:CAK31432.1; GSPDB:GT

C/Genetics:

A/Gene: xciC

C/Superfamily: 4-coumarate-CoA ligase; acetate-CoA ligase homology

Query Match 9.7%; Score 298; DB 2; Length 476;
Best Local Similarity 26.4%; Pred. No. 1.5e-15;
Matches 135; Conservative 61; Mismatches 196; Indels 120; Gaps 24;

```
QY 44 SYAEMFNVAIAQSLPLPGLSAERPPLIVSGNDLHLQLAFAMTAAGIPYCPVSPAYSL 103
DB 38 SVABRVGARLVAVLATP---TASTVLAIR-----GGLIAGVFVVPVADIGV 82
QY 104 LSQDLAKLRITV-----GLQGLVFAADAPQRAITLITLDDVPAITFGEHAGRT 157
DB 83 VER-----RHMLTDSGAQAWLIGEP--SADSA-----PDGLPHI-----P 115
QY 158 VSFDSLLEQCGGIEADNAAFAATGPDTIAKFLFTSGSTGLPKAVPTTORMLCANQWMLQ 217
DB 116 VQIDA-----RSMRYPESEPDATAMVYISGTTGPKGVLSRAIAYDADALAA 167
QY 218 FVYFGESEPVLYDMLPMNHTFG-----GSHNIGIYLVNGGTYVDDGKPTAOGAETLR 271
DB 168 WQMTATDD-VLYNGGLPLFVHGLVGLGSLR-----GNFVHTGKPTTAYVAAQ-- 216
QY 272 NLSEISPTAYLTVPKGMBELVGLERDSTLRERFFARMKLFPAAGLSGICIDRLDRA 331
DB 217 -CSEAGSGLYFGVPTWRSRLVA---DEAV-ARALPRLVLSGSASLPPVPEDRLAHLT 270
QY 332 EQHGERIRIMMAGLMTETAPRSCFTTGTGLSAGYGLPAPGCEVLYDMLPMNHTFG 384
DB 271 GHRPIER-----YGSTESLITLSTLADGERRAGWGLPLAGVQTRLVDSGGPVYDGE 324
QY 385 LGRF--HGPHVMSGYWRAPQNAQAPEDEGGYSCGDAIKLADPADPQKGLMFGRIAE 442
DB 325 TVRLQVRSPTMGTYLNREATAEAFDEDEGWTGTD-VAVVDSGGHNR---IVGRESVD 380
QY 443 FKLSGTVFVSGLPRTAIVLEGGSYVLDV-VVAAPDEEGLGLVPRLLDCAALSGLGKE 501
DB 381 LKILGVRIGAGRI--EMALLGHPDVRVAVVGLPD-EDLGRIVAFV-----VGAE 429
QY 502 ASDAEVLAS-----EPVRAWFADMLDKR 523
DB 430 ALDADELINVAOQLSIHKRPREVRVVDALPR 461
```

RESULT 6

AG0066

probable AMP-binding enzyme-family protein YPO0537 [imported] - *Yersinia pestis* (strain CC/Species: *Yersinia pestis*

C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004

C/Accession: AG0066

R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; f

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, f

Nature 413, 523-527, 2001

A/Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AG0066

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-601 <RUR>

A/Cross-references: UNIPROT:Q82IG5; GB:AL590842; PIDN:CAK89394.1; PID:gl5978630; GSPDB:GT

Qy	424	ADPADPQKGLMFDR	IAEDFTLSSGVFVS	VEPLTRAVLEGGSYL	VDVVAAPDEECGL	483
Db	425	AVADDPGRGF	FEDGMAENFKL	QGTGTWVA	VGFLRAQLVMFAGLLRDV	VTGENSEAEILGA 484
Qy	484	LVFPRLLDCRL	SLGSKSA	DAEVLASEPVR	WPFADWLKRLNREATGNASR	ITMWGLDT 543
Db	485	LVFPRIPALRE	LVRSQHLSDAE	ITIRHPSVRAQI	VAKLSSAHQKQASGASR	VMRLVME 544
Qy	544	PPSIDKGEVTDK	GSINORAVLQ	WMSAKVDALY	575	
Db	545	ALREKGEVTDK	GSINORAVL	LHREKLEVESLY	576	

RESULT 2

long-chain-fatty-acid-CoA-ligase Atu1416 [imported] - Agrobacterium tumefaciens (strain C:/Species: Agrobacterium tumefaciens
C:/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:/Accession: AM2750
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monk, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, P.; Grant, C.; Guenther, D.; Kutuyavin, T.; Levy, R.; Li, M.; McCell
i, K.P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:/Reference number: AB2577; MUID:21608550; PMID:11743193
A:/Accession: AM2750
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-611 <KUR>
A:/Cross-references: UNIPROT:Q8UFI1; GB:AE006688; PIDD:AA142422.1; PID:g17739834; GSPDB:G
C:/Experimental source: strain C58 (Dupont)
C:/Genetics:
A:/Gene: Atu1416
A:/Map position: circular chromosome

Query Match	Similarity	42.0%	Score 1294	DB 2	Length 611
Beet	Local	46.5%	Pred. No. 5.6e-94		
Matches	266	Conservative	84	Mismatches	220
				Indels	2
				Gaps	2
Qy	5	EALLPFRRLIEREHNMAKTRPEQTCVAARANGSEWRISVAEKPFNNVRAIAGSLPYGL	64		
Db	26	PLPGVPGKLMERLLHWCSRABERTWMAADRGREBWRVSVYAEALDKIRRIQGLDHL	85		
Qy	65	SAERPLIVSGNDSEHLDAFGAMTAGIPLYCPVSPAYSLSQDLAKRHYVGLLQPGLVF	124		
Db	86	SVERPLVLASNSIEHALMVLAAQHVGLASAIITPVA-TSADITLKADIRGQITTPQVVF	144		
Qy	125	AADAPFORAIEITLLPDDVPAIFTRGELAGR-RVYSPSLSEPOGGEIADNAFAATPDT	183		
Db	145	AEDATPFRRALGEVFDOSTPLVGLRNLPEDRSNTHFETLTLETPEVAVDRAFAVGPDT	204		
Qy	184	IAKLFTFSGSFKLREKAVPTTORMLCANQOMLLQTFPVNFGSEPPVLVDMLEPNHHTFGGSHN	243		
Db	205	VAKELFTSGITGSPRAVIGTQRMCSNQEMADCGYFREBPVVVVMAPNNHNAANKV	264		
Qy	244	IGILVYNGGTYYLDDGKPTAAGPATTLENLSEISPTATLVYPKMBELVGLERDSTLRE	303		
Db	265	FNVLVYNGGTYYIRGKRSPAOIGOTLDNLADISPTWYFNVYPAGHEMLVQVMRDEALCR	324		
Qy	304	RFPARMKLFPPAAAGLSGCIWDRIDRAVYAEHQGSEIRIRMAAGLMTETAPSTPTTGPLSM	363		
Db	325	SFFDDLKMLMAGAGMAQHTDALTELSTMAVTVGAHVLGAGLSGETEPAFSLFTCEBQDK	384		
Qy	364	AGVIGLPARGCEVTKLPVYDGKLEGRFHGPHMWSGWRAPBEONAAQFBEEGYCGSGDAIKL	423		
Db	385	PGNIGIPARGVTKLVPEFDGRYTELRKQPNITTPGWRMRGELTAAAFBEGRYRIGDIYKF	444		
Qy	424	ADPADPOKGLMFQRIADPEFKLSSGVFVSGPLRTRAVLBCGSYVLDVVAAPRECLG	483		
Db	445	AVADDPRGGEFTDGRMAENFTLQGTGTWAVAPRLQNLNMAGILIRDAVITGEENRAEBGA	504		

```

QY      484  LVPEPLDORALSGCKEASDEAVTASBPVAMPADMLKRLNREATGNSASIVVGLDT 5433
Db      505  LVVPEPLPALRELVRSGCHLSDAEITIRHRSVRAQIVAKLSAHQKQASGSASRVRILVMD 5644
QY      544  PPSIDKGEVTDKGSINQRAVILQWRSKADYALY 575
Db      565  ALRFEKGEVTDKGSINQRAVILHRLRELVEESLY 596

```

RESULT 3

feruloyl-CoA synthetase [imported] - *Caulobacter crescentus*
 C.Species: *Caulobacter crescentus*
 C.Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
 C.Accession: F87297
 R.Nieman, W.C.; Feldhlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J. I.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolome
 N. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A.Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A.Reference number: AB87249; MUID:21173698; PMID:11259647
 A.Accession: F87297
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-596 <STO>
 A.Cross-references: UNIPROT:Q9AB43; GB:AE005673; NID:913421550; PIDN:AAK22378.1; GSPDB:G
 C:Genetics:
 A:Gene: CC0391

Query Match	32.5%	Score 1002.5	DB 2	Length 596
Best Local Similarity	40.5%	Pred. No. 5.4e-71		
Matches 227	Conservative 81	Mismatches 227	Indels 25	Gaps 7
QY	18	LEHNAKTEPEQTCVAAARAANGEMRISYAEFMFNHVAIQSLPYGLSAERPLIVSGND	77	
DB	49	LARMAAQAQPDRIWVLELRDEQG--WRTVTYADAKAQVEALGGGLGRLRGCPILLIIRANG	107	
QY	78	LEHLOLAGAMVAGIPIYCPVSPAYSLISODLAKLRITVGLLOGLVPAADAAFPQAIET	137	
DB	108	VDHARIAAANAASLGPAPAPVSPQYGLKGADLRLOHAHVEKLKAAVYADDAEAAGDLAA	167	
QY	138	ILPDDVPAIFRTGELAGRTVSPFDSLLEQPGGIEADNAFAATGPDTIAKFLFTSGTKLP	197	
DB	168	PPLAGLPVVVSNAPRG--DVAFAELIKSP-----PLAPQAQPDVAKILLISGSGTKP	219	
QY	198	KAVPTTQRMLCANQOMLQTFPFVPGEEBPVLVDMLPMNHTFGSGHNIGVLYNGCTYYLD	257	
DB	220	KAVVCTHANIAALNAQAIEACVA--DDPPLVLVSAPWSHSLGANNAIIMHVLHRCGTIYID	277	
QY	258	DGKPTAQAQPAETLRMLSEISPTAYLTVPRKGMBELVQALBEDSTLRERFPARMKLPFAA	317	
DB	278	AGQAPRGRRGETVRLNISEVATTYHNHNPAGMANFVELELDEALAAFFEKVRVLQDGA	337	
QY	318	GLSGGIMPLRLDRAVQOHCGERIRMMAGLGMTETAPSCPTFTGCLSMAGYIGLPAPGCEVK	377	
DB	338	SMAQGISLDRVQAVAVARTGERITTPAAGYGATETGFPACINIMHINARSQWGLPPIRGAVK	397	
QY	378	LVPV--DGKLEGRFHGPHVMGSGYWRAPBEONQAQFDEEGYICSGDAIKLADPADPQKLMF	435	
DB	398	LVPAGEGGGFETIRVKGPGVSPGLDQPEATAQAFDEDDGFRLDPAALADBEDPAAGIVF	457	
QY	436	DGRIADPFLSGSVSVGLRTRAVLBEGSVYLDVYVAAAPDEBCGLVLPFRLDQAL	495	
DB	458	DGRLVENFTLAAAGTIVAGALRYAAVASAIGGVSDAVVCEGEGBGVGLMF--LDKAV	514	
QY	496	SGLGKASDAEVLASBPVWAFADWLKLRNREATGNASRIMWGLLDTPTSIDKEVTDY	555	
DB	515	ERLG-----SPEAVAAIWEALSRMNAAKGGGRIARALIIDGADDAASGETLTDK	565	
QY	556	GSINQRAVLQWRSAXVDALY	575	
DB	566	GYINQALARDRRPKELERLF	585	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2005, 23:20:35 ; Search time 39 Seconds
(without alignments)
1453.120 Million cell updates/sec

Title: US-09-750-986D-30

Perfect score: 3082
Sequence: 1 MRSLEALPPGRILRLERH.....KVDALYRGEDQSMRLDEATL 589

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1294	42.0	591	2 G97531	feruloyl-CoA synth
2	1294	42.0	611	2 AH2750	long-chain-fatty-a
3	1002.5	32.5	596	2 F87297	feruloyl-CoA synth
4	972	31.5	621	2 H87662	feruloyl-CoA synth
5	298	9.7	476	2 E87040	acyl-CoA synthase
6	289	9.4	601	2 AG0066	probable AMP-bind
7	286	9.3	607	2 H64041	probable long-chain
8	283.5	9.2	601	2 H82069	probable long-chain
9	273	8.9	672	2 T07944	probable long-chain
10	270	8.8	721	2 JC7557	lipidosis - mouse
11	265	8.6	601	2 H69274	probable long-chain
12	261	8.5	536	2 E85558	2,3-dihydroxybenzo
13	260	8.4	536	2 A99708	2,3-dihydroxybenzo
14	258	8.4	566	2 H82132	long-chain-fatty-a
15	257.5	8.4	600	2 E70937	probable fadD15 -
16	257	8.3	508	2 H95271	probable long-chain
17	256	8.3	536	1 SYCEB	2,3-dihydroxybenzo
18	255.5	8.2	687	2 T27421	hypothetical prote
19	254	8.2	656	2 C71361	probable long-chain
20	251	8.1	558	2 T34849	probable acetyl-co
21	250.5	8.1	683	2 AC2256	hypothetical prote
22	250	8.1	473	2 G70607	probable fadD36 pr
23	247	8.0	645	2 H70173	long-chain-fatty-a
24	246	8.0	699	2 A36275	long-chain-fatty-a
25	244	7.9	638	1 JX0202	probable non-ribos
26	244	7.9	4342	2 H83343	brain-specific lon
27	243	7.9	722	2 JC7970	probable long-chain
28	242.5	7.9	600	2 A87020	probable acid-CoA
29	242	7.9	509	2 C69491	probable acid-CoA

30	242	7.9	647	2 T30892	hypothetical prote
31	240.5	7.8	550	1 A26772	Proteinus-luciferin
32	239.5	7.8	598	2 T35513	probable long-chain
33	238.5	7.7	696	2 S74994	long-chain-fatty-a
34	237	7.7	547	2 A70551	probable acid-CoA
35	236	7.7	827	2 S72741	probable polyketid
36	236	7.7	1188	2 D86969	probable acyl-CoA
37	235.5	7.6	530	2 B87369	long-chain-fatty-a
38	235.5	7.6	612	2 T35430	probable long-chain
39	235.5	7.6	720	2 E84456	probable acyl-CoA
40	235	7.6	584	2 C75364	probable long-chain
41	233	7.6	525	2 AG0232	Yersiniaabactin sid
42	233	7.6	525	2 T17443	salicyl-AMP ligase
43	232.5	7.5	823	2 D70386	probable polyketid
44	232	7.5	555	2 C83444	probable AMP-bind
45	232	7.5	650	2 H84356	long-chain fatty-a

ALIGNMENTS

RESULT 1

G97531
feruloyl-CoA synthetase (A1238746) [imported] - Agrobacterium tumefaciens (strain C58, C
C/Species: Agrobacterium tumefaciens
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C/Accession: G97531
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: G97531
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-591 <KIR>
A/Cross-references: UNIPROT:Q8UFI1; GB:AE007869; P1DN:AAK87208.1; P1D:G15156488; GSPDB:G
C/Genetics:
A/Gene: AGR_C_2614
A/Map position: circular chromosome

Query Match	Score	DB 2;	Length	591;
Best Local Similarity	46.5%	Pred. No. 5,4e-94;		
Matches 266;	Conservative 84;	Mismatches 220;	Indels 2;	Gaps 2;
QY	5	EAALLPPGRILRLERHMAKTRPEOTCVARARANGEMRIRISYAEWFRHRAIAGSLPYGL	64	
DB	6	DPLGPIYQKLNERRILHMCRAPERTWADROGREPMRVRVYALDKIRIKGFLDHD	65	
QY	65	SABRPILIVSGNDLEHLQAFGMYAGIPYCPVPSPAYSLSDILAKLRHIVGLIOPGLVF	124	
DB	66	SVERPLIVSSENGIEHLAVLAAQHVGIASAITPAVA-TSADUTKLADIRGQTTPGMVF	124	
QY	125	AADAAFPQRAIETILPDVDAITFRGELAGR-RTVSGDSILBORGGEIADNAPATPDT	183	
DB	125	AEATATPRRALGCVFDGFLVGLRLNLPEDRSNTFHETLTLETPTEAVVRAAPAVPDT	184	
QY	184	IAKFLFTSGSTKLPKAVPTTQRMICANQOMLTQTPPVGGEPRVLVDMLEPMNTHFGSSHN	243	
DB	185	VAKFLFTSGTSPKAVIQTQRMICSNQEMTADCYGTFREPRVVDMAAMNHTAAGNKV	244	
QY	244	IGILVNGGTYYLDDGKPTAGPAETLRNLSEISPTAYLTVPKGWEELVGALESDSTLRE	303	
DB	245	FNILVYNGGTYYLDRGKPSPAQIGQTLNDLRDISPTYYFNVVAGHEMLVQAMKDEALCR	304	
QY	304	RFPARMKLFFPAAAGSGIGIWDRLDRAVEQHCGERIRMAAGLMTETAPBCTTTTGLSM	363	
DB	305	SFPRDLKMLMYAGAGQAQHTMDALTELISMAVGVHVAVMGSLGSLTEPAAPSLCTBEPDOK	364	
QY	364	AGYIGLPAPGCEVKLVVDGKLEGRPHGPHVMSGYWRAPQNOAPDEEGYSGDAIKL	423	
DB	365	PGNITGLPAGCVTKLVPDFDGRYELRLKGPNTITGGMNGBLTAAPDEEGFYRIQDTVKF	424	


```
QY 470 AspValValAlaAlaProAspArgLysLeuValPheProArgLeu 489
Db 1420 CAATAATGATGTTGGTGAAGGAGAAATCCCTTCGTGATTTGGCGCGTTC 1479
QY 490 LeuAspCysArgAla-----LeuSerGlyLeuGlySerGlyAlaSerAspAlaGlu 506
Db 1480 GCAGATGTAAAGCTGGTGTGAGATCAAGCAATT--CCGTATACGACAAAGAGAA 1536
QY 507 ValLeuAlaSerLeuProValArgAlaTrpPheAlaSerTrpLeuLysArg 526
Db 1537 ATGGTTGCATGACGACGCGTATCGAAATTCAGAAAGATTGACGAATCAT--- 1593
QY 527 GluAlaThrGlyAsnAlaSerArgLysMetTrpValGlyLeuAspTrpProSer 546
Db 1594 GAAACCTTGCACAGTATGAACGTGAAGAAATTCATTTAATGCCAAGCCATTACG 1653
QY 547 IleAspLysGlyGluValThrAspLysGlySerLysAsnAlaValLeuGlnTrp 566
Db 1654 ATAGAGAGCGGAGAAAGACACCGACATTGAAATTTGAACGTATGATTCATTAGCAGAA 1713
QY 567 ArgSerAlaLysValAspAlaLeuTyr 575
Db 1714 TGCCAGGCGCAGATAGAGAAATGAT 1740

RESULT 15
US-10-470-565-1
; Sequence 1, Application US/10470565
; Publication No. US20040126870A1
; GENERAL INFORMATION:
; APPLICANT: Societe des Produits Nestle S.A.
; TITLE OF INVENTION: NCC2705 - the genome of a Bifidobacterium
; FILE REFERENCE: 80290/WO
; CURRENT APPLICATION NUMBER: US/10/470,565
; PRIOR FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: EP 01102050.0
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2256646
; TYPE: DNA
; ORGANISM: Bifidobacterium longum
US-10-470-565-1

Alignment Scores:
Pred. No.: 3 77e-20 length: 2256646
Score: 314.50 Matches: 146
Percent Similarity: 38.70% Conservative: 87
Best Local Similarity: 24.25% Mismatches: 235
Query Match: 10.20% Indels: 134
DB: 17 Gaps: 23

US-09-750-986D-30 (1-589) x US-10-470-565-1 (1-2256646)
QY 14 IleLeuGluArgLeuGlnHisTrpAlaLysThrArgProGluGlnThrCysValAlaAla 33
Db 354867 CTACACGATCGCCT-----GCCCGACCGCGCGGACACGGAATCGCGCCAG 354917
QY 34 ArgAlaAlaAsnGlyLysTrpArgArgIleSerTrpAlaGluMetPheHisAsnValArg 53
Db 354918 AAAATCGCGCCCGCGCGCTGCAGAAAGCTACCAACCGCGCAATTTCCACGAGCGGTGTC 354977
QY 54 AlaIleAlaGlnSerLeuLeuProTyrGlyLeuSerAlaGluLysArgProLeuLeuVal 73
Db 354978 TCGCGCCGCAAGGCGCTGATCGCGCTGCGCATCCCAAGGCGCAGCGGTGACCATCTTT 355037
QY 74 SerGlyAsnAspLeuGlnHisLeuGlnLeuAlaPheGlyAlaMetTrpAlaGlyIlePro 93
Db 355038 TCTCTCCACCGCGTCTAGAAATGGGGCATTTCTAGATTTCACATCGCGCGGTGGGCGCGTA 355097
QY 94 TyrCysProVal-----SerProAlaTyrSerLeuLeuSerGlnAsp 107
Db 355098 AGCGTCCGATCTATGACACCGAATCGCGCCCGCAGCGCCACGCGCATGTAAGATTTCG 355157
```

```
QY 108 LeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAsp 127
Db 355158 GTGCTAAG-----CTGGCTTTGCCGACAT 355184
QY 128 AlaAlaProPheGlnArg-----AlaIleGlnTrpIle 138
Db 355185 CGCGAGCGCTTCACCGCGCTCGACTCGTCACAGACCATGCCCTGCTTGAAGCAAT 355244
QY 139 LeuProAspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArg---Thr 157
Db 355245 CTCATGATCGAGGCAACGCGCTG-----GGCGCTTGAAGCGCTTAGCGGTAC 355295
QY 158 ValSerPheAspSerLeuGlnGlnProGlyIleGluAlaAspAsnAlaPheAla 177
Db 355296 GTTTCGACGAGAACTGCACGAGCGCGGTGCCACCGCTGCCACGAGCAGC----- 355346
QY 178 AlaThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuPro 197
Db 355347 -----CTGGCCACCATCGTCTACCTTCGGGCTTCACCGCCACCG 355388
QY 198 LysAlaValProThrThrGlnArg-----MetLeuCysAlaAsnGln 212
Db 355389 AAGGTGCGCACTGACCCACAGAACTTGTCTCATCACTCCGCTCGCAGCGC 355448
QY 213 MetLeuLeuGlnThrPheProValPheGlyGluGluProProValLeuValAspTrpLeu 232
Db 355449 TTG-----CATGAAGTGTGCTTGAAGACCATCGCGCGCTGCTTCCTG 355496
QY 233 ProTrpAsnHisThrPheGlyGlySerHisAsnIleGlyIleValLeuTyrAsnGly 252
Db 355497 CCGCTGGCCCATCGTTCGCGCATTCATCCAGTACGCTTCCTCCGCTCCAGTACGGC 355556
QY 253 ThrTyr---TyrLeuAspAspGlyLysProThrAlaGlnIlePheAlaGluThrLeuArg 271
Db 355557 GTGGCTGTTATCTGCGGACACCA-----ACCTGCTGCGC 355595
QY 272 AsnLeuSerGluLysSerProThrAlaTyrLeuThrValProLysGlyTrpGluGluLeu 291
Db 355596 GACCTGCTTCTGTTGAACCGACCTACCTGCTGGGGTGCAGCGGTTCGAGAAAGTA 355655
QY 292 ValGlyAlaLeuGluArgAspSerThrLeuArgGluArgPheAlaArgMetLysLeu 311
Db 355656 TACACGCGCCCTCGCACCAAGCGCGT-----GCCGCTGAAAGGGC 355697
QY 312 PhePhePheAlaAlaAlaGlyLeuSerGlnGlyIleTrpAspArgLeuAspArgValAla 331
Db 355698 CGCCTGTTGTCAAGCGCGCGAGCGCGCGCTGTGAGCCGCAAGGACGAGCGCGC 355757
QY 332 GluGlnHis----- 334
Db 355758 GAGCAGACACGTTTGCGCAAAATCGCGAGCGGCCAATGACGAACCGTCTACCGC 355817
QY 335 -----CysGlyGluArg----- 338
Db 355818 ACGGTGCGGCTGCTGCGCCGGAAGATCAAGTAGTGCGCGCGCGCACCATG 355877
QY 339 -----IleArgMetAlaGlyLeuGly 346
Db 355878 TCGCTGANTGGCGCATTTCTAACAAGCATGCACTGCGAGTACGAGGTATGGC 355937
QY 347 MetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyr 366
Db 355938 ATGACGAAACCGCCCGCGCTTCGCGCGCGCGGTGACGACAAATGATGATCGGAC 355997
QY 367 IleGlyLeuProAlaProGlyCysGluValLysLeuValProValAspGlyLysLeuGlu 386
Db 355998 GTAGGCGAGCGCGCGCGCTCTCGATTGCG---ATTCCGACGAGCGAAGCTGCGAG 356054
QY 387 GlyArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAla 406
Db 356055 GTCAAG-----GCCCGAAGCTGTTCCGCGGTACCAATCTCGCGAGAAAGCGCC 356108
```

QY 528 aThrGlyAsnAlaSerArg-----IleMetTrpValGlyLeuLeuAspThrPro 545
 Db 6728952 ----GGTGGCTCCAGGCTGCGAGACCACTAAGAACTTCCCTCTCCGCGGACACT 6729007
 QY 545 oSerIleAspGlyGlyValThrAspGlySerIleAsnGlnArgAlaValLeuG1 565
 Db 6729008 CGACATCGAACACGCGGAGATGATCCCGAGCTCAAGCTGAGAGCGCCGCTCGCAACG 6729067
 QY 565 nTrpArgSerAlaValAspAlaLeuThrArgGlyValAspGlnSer 581
 Db 6729068 CGAGTCAAGACATCTGATCGACGATGTCGCGGGTCCGCGGAGGCT 6729116
 RESULT 14
 US-10-369-493-34073 Application US/10369493
 / Sequence 34073.
 / Publication No. US20030233675A1
 / GENERAL INFORMATION:
 / APPLICANT: Cao, Yongwei
 / APPLICANT: Hinkle, Gregory J.
 / APPLICANT: Slater, Steven C.
 / APPLICANT: Goldman, Barry S.
 / APPLICANT: Chen, Xianfeng
 / TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 / FILE REFERENCE: 38-10(52052)B
 / CURRENT APPLICATION NUMBER: US/10/369,493
 / PRIOR FILING DATE: 2003-02-28
 / PRIOR APPLICATION NUMBER: US 60/360,039
 / NUMBER OF SEQ ID NOS: 47374
 / SEQ ID NO 34073
 / LENGTH: 1746
 / TYPE: DNA
 / ORGANISM: Cytophaga hutchinsonii
 US-10-369-493-34073
 Alignment Scores:
 Pred. No.: 1,4e-25 Length: 1746
 Score: 324.00 Matches: 144
 Percent Similarity: 40.23% Conservative: 101
 Best Local Similarity: 23.65% Mismatches: 264
 Query Match: 10.51% Indels: 100
 Gaps: 26
 US-09-750-986d-30 (1-589) x US-10-369-493-34073 (1-1746)
 QY 26 ProGlnGlnThrCysValAlaAlaArgAlaAlaAsnGlyLutTrpArgAlaSerTyr 45
 Db 37 CCCAAACAAAGATGCACTGCGCAACCAA--GAGAAATGCGACCTGATTAATTCCTT 93
 QY 46 AlaGluMetPheIleAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeuSer 65
 Db 94 GTGCGGTTCAGAAATTATGATTAAGTAAGTATTGGCTTTTAAATCGGGGTTTACAA 153
 QY 66 AlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeuAlaPhe 85
 Db 154 AAAGACGATTAAGTACCATTTATTCACCAACCGCTCCGAAATGGAATTTATGATCTG 213
 QY 86 GlyAlaMetCysTrpAlaGlyTyrProTyrCysProValSerProAlaTyrSerLeuLeuSer 105
 Db 214 GGACTCCAAACAGATAGACGCGGTGAGTGTATCCGATTTATTCG-----ACACTTACGCTG 267
 QY 106 GlnAspLeuAlaValSerArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAla 125
 Db 268 GAAGAT-----TACCGTATATTTTGTATGATGACCGCGAAGTTAAATGTTTGTG 318
 QY 126 AlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspVal----Pro 144
 Db 319 TATGACGAACAATATTATTTGCAAAAGCAACGAAAGCGTAAACGGTCCGAAACAATCCT 378
 QY 145 AlaIlePheThrArgGlyLeuValaGlyArgArgThrValSerPheAspSerLeuLeu 164
 Db 379 GTAAATATATACCTTGACGAGATGTGAACGAGCAAAAGCTGGAAGATATACGCGCACTG 428

QY 165 Glu-----GlnProGlyGlyIleGluAlaAspAsnAlaPheAlaIleThrGlyProAsp 182
 Db 439 GCAGACGCCAACCGGCTTCGATG---CTTGATCCATATTAAGCAATATTAATTAACGAC 495
 QY 183 ThrIleAlaValPheLeuPheThrSerGlySerThrIleLeuProValAlaValProThr 202
 Db 496 GATCTGCTTAACCTGATTTATATACGTCGGAACCTACAGTAAGCCCAAGGGTGTATGCTT 555
 QY 203 ThrGlnArgMetLeuCysAlaAsnGlnGlnMetLeuLeuGlnThrPheProValPheGly 222
 Db 556 ACACATTAACAATATTAATTAAGCAATGTAATCGACGCAATTCGCTGAATGCGCGTT----- 609
 QY 223 GlnGluProProValLeuValAspTrpLeuProTyrAsnHisIleThrPheGlySerHis 242
 Db 610 AATCAACATCATTAAGCACTTAAGTTTATACCTGTTGCGACGATAT-----GAA 660
 QY 243 AsnIleGlyIleValLeuTyr-----AsnGlyGlyThrTyrTrpLeuAspAspGlyVal 260
 Db 661 CGCATGTGTATCTGTATACCTGCTCAGGGGGTATCTATATATATGACGAAAGTACA 720
 QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
 Db 721 GATACAGTTGCG-----GAAATTTAAAGAAAGTGCAGCGCACATG 762
 QY 281 TyrLeuThrValProValGlyTyrGlu----- 289
 Db 763 TTGTTCAGGTGCGCAAGCTGCTTGAAAAAGTATACGATAAATTTGCTTAAGGCGCT 822
 QY 290 GluLeuValGly----- 297
 Db 823 GAATCTTACGGGCAATTAATAATTCCTCTATATGGGCACTGAGCTGGGGTTAAATAT 882
 QY 298 AspSerThrLeuArgGluArgPhePheAlaArgMetLeu-----Phe 312
 Db 883 GATATGCAAGAAATCAAGGCTTCTTTTAAATTCATCACTGAAGCTTCCATTAATAATA 942
 QY 313 PhePheAla-----AlaAlaGly-----LeuSerGlnGly 322
 Db 943 ATCTTACGTAATATGCGAAGACATTTGGGTGGAAATGATTAATCCCTGTTTACGCGGT 1002
 QY 323 -----IleTrpAspArgLeuAspArgValAlaGlnGlnHisCysGlyGlyValArg 340
 Db 1003 GCTGCTTACAGCTGCTTTGGCGCGT-----TTCGCGCGCGCGAGATTCG 1053
 QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPhe----- 356
 Db 1054 GTGATGAAAGTTTACGTTTAAACAGACGCTGCGCGGTATTCGCTTAACAGATTCAT 1113
 QY 357 -----ThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeu 369
 Db 1114 CCGAGTAAACACGAAATCACTACTGTGGTCCGCT----- 1149
 QY 370 ProAlaProGlyCysGluValIleValLeuValProValAspGlyValLeuGluArgPhe 389
 Db 1150 -----TTGCGCGGAGTAAAGTAAGATCGCGGAA--GACGCGAAATTCGACGCGC 1200
 QY 390 HisGlyProHisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAlaGlnAlaPhe 409
 Db 1201 ---GGACCGCAATATTAAGAAAGTACTCAAAACGCGCGGATTTTAAACGCGCAAGCAATT 1257
 QY 410 AspGluGluGlyTyrTyrCysSerGlyAspAlaIleIleLeuValAspProAlaAspPro 429
 Db 1258 GACAGTGAAGGCTGTTTACACAGAGATATCGGTCACTTTGTAGAA----- 1305
 QY 430 GlnValGlyLeuMetPheAspGlyArgIleAlaGluAspPheValLeuSerGlyVal 449
 Db 1306 AATAAATATTTAAAGATCACCGACCGTAAAGAAATGTTTAAACCTCGGGGTGTAAG 1365
 QY 450 PheValSerValGlyProLeuArgThrArgAlaValLeuGluGlyGlySerTyrValLeu 469
 Db 1366 TACATTGCA-----CCGACGCTGTTGAATAATTAATTAAGAGTCCGCTACATTGAA 1419

Oy	101	TyrSerLeuLeuSerGlnAspLeuAlaIysLeuAaGhIstIleValGlyLeuLeuGlnPro	120
Db	253	TACGGCAATTATCTTGAGAGATG-----ATGACTTACTGTCTGTCAATTGGCAAAACC	306
Oy	121	GlyLeuValPheAla-----	125
Db	307	GCAATTAGTCATCCGAAAGGATACCTTGACAGACACTGGACAGCATGATGACCCGTGG	366
Oy	126	-----AlaAspAlaIleProPhe-----GlnArgAla	134
Db	367	CAGGTGATGACACTGGAACAGGCCGACCTTTCCCTGATTTGGCGATTTGACACGGGAA	426
Oy	135	IleGlnThrIleLeuProAspAspValProAlaIlePheThrArgGlyLeuLeuValagly	154
Db	427	ACTGACCTGATCAGCCCGGATTAAGCCGTCAGTCATGAAGATATCGAGAG--GCGTTA	483
Oy	155	ArgArgThrValSerPheAspSerLeuLeuGlnInProGlyGlyIleGlnAlaAspAsn	174
Db	484	CTCAGGGGTATATCCAGACAGATTTACAGCCACGCG-----GCAACACAG	528
Oy	175	AlaPheAlaIleThrGlyProAspThrIleAlaIysPheLeuPheThrSerGlySerThr	194
Db	529	GCATTTCTGT-----TTGATGTTTACTTCTGTGTTCCAGT	561
Oy	195	LysLeuProIlyValAlaValProThrThrGlnArgMetLeuCyAlaIAsnGlnInMetLeu	214
Db	562	GGTAAACCGAAAGGTGTGTGCTCAGTCAGAAACAATATT---CTGTACACAGACAGAGCT	618
Oy	215	LeuGlnThrPheProValPheGlyValGlnIupProProValLeuValAspThrPheProThr	234
Db	619	CTGCAACAATTATGGCAATTC---GATGCCCGCAATCTTTTCTAGTTACTGCTGCCGTGG	675
Oy	235	AsnIsthrPheGlyGlySerHisAsnIleGlyIleValLeuTyAsnGlyGlyIlyThrTy	254
Db	676	CATCACACCTTGGTGATTTATTGAAATAACAGCGCATATTTGACAGCGCATGTTA	735
Oy	255	TyrLeuAspAspGlyLysPProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSer	274
Db	736	CATATGTGATTAATAGCCGG-----GGCATGATATTGACCGGATGTGACATATCTGAGT	789
Oy	275	GluIleSerProThrAlaTyrLeuThrValProIlyGlyTyrGluGluLeuValGlyAla	294
Db	790	CAGTTAAACCGACACGTTTTTTCAGTGTATCCCAAGATACATGCACAGTGTGCCAACA	849
Oy	295	LeuGluArgAspSerThrLeuArgGluArgPhePheAlaArgMetLysLeuPhePhe	314
Db	850	ATGAAACATTCGAGCGCTTCCACGGGACAGCGACCGACACTGCATGATCTTCTCC	909
Oy	315	AlaIleAlaGlyLeuSerGlnGlyIleThrAspArgLeuAspArgValAlaGluGlnHis	334
Db	910	GCTGCTCTCAAAATGCTCCCGGATATCGAAGCTATTTT-----AAGCAT	954
Oy	335	CysGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCys	354
Db	955	TGTGGT-----ATCTACGTTGCTGAGGGCTGGGGGCTGACGGAACACAGCCA---TGC	1009
Oy	355	ThrPheThrThrGlyProLeuSerMetAlaGlyTy-----IleGlyLeu	369
Db	1006	CTGACACTTAAAGGCCG-----GGCTATGAAAGAACACMACAGCGGTAGCCGC	1053
Oy	370	ProAlaProGlyCyGluValIlyLeuValProValAspGlyLysLeuGluGlyArgPhe	389
Db	1054	CCTTTATCCCAATGTCATTTCTCAGTTATGATGAAGATAGCGGGAGATTTTGGCAAAA---	1110
Oy	390	HisGlyProHisValMetSerGlyTyrThrArgAlaProGluGlnAsnAlaGlnAlaPhe	409
Db	1111	---GGGCGGAATGATGATCTGGGCTATATATGCGCATGAAGCCGCTAACACAGGTTGCTTC	1166
Oy	410	AspArgGluGlyTyrTyCysSerGlyAspAlaIleIysLeuAlaAspProAlaAspPro	429
Db	1168	---AGGACGGGATGGTTTCAAGAACCGGCATCTCGGCAAGATGATC-----	1208

QY	430	GLINYSGLY---	LeuMetPheAspGlyArgIleAlaGluAspPheIysLeuSerSerGly	44
Db	1210	---AAGGGTGA	CTTGTCGTTTCGGGGCGGTTTGATTCAGTCAAAAAGCTGTAAATGGC	12
QY	449	ValPheValSerValGlyProLeuAspGlyThrArgAlaValLeuGluGlySerTyPval	46	
Db	1267	GAAGAAGTATCCAGCAGTGCATTGAA---	CAAAAGTCTGTTTACGAAACCGATATTATTC	13
QY	469	LeuAspValValValAlaAlaIaProAspArgGluCysLeuGlyLeuLeuValPhe-----	48	
Db	1324	AACATGCCATTATATGATATTGAACAACGCCCTACGCTACCGCTTATTTGTTATTAAAC	13	
QY	487	ProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyIysGluAlaSerAspAlaGlu	50	
Db	1384	CTTGAGCTG-----	TGTCCGCCAGCGATTGTG-----GAGATTTCACA	14
QY	507	ValLeuAlaSerGluProValaArgAlaIaTrpPheAlaAspTrpLeuIysArgLeuAsnArg	52	
Db	1423	CTGCACACATCATCCCTTGCCAGACAGCATATTCCTGAGGTGTATCAACGCACTAAATCAG	14	
QY	527	GluAlaIaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThrProProSer	55	
Db	1483	CAGGTGAGAGATAGTGTGTTTAATGATTCGGGTGGCGGTACTGAACCTCAGCTGT	15	
QY	547	IleAspIysGlyGluValIleThrAspIysGlySerIleAsnGlnArgAlaValLeuGlnTrp	56	
Db	1543	CTGAGAAAGGGGGAAGTCAACCCGTCATTTAAAGTGTCCAGCGTACTGTACCGCAGAAAT	16	
QY	567	ArgSerAlaIysValAspAlaIaLeuTyx	57	
Db	1603	TATCAGCGTGTGCGTGAATGCCCTTATAC	1629	
RESULT 12				
	US-10-156-761-5544			
	; Sequence 5544, Application US/10156761			
	; Publication No. US20030119018A1			
	GENERAL INFORMATION:			
	APPLICANT: OMURA, SATOSHI			
	APPLICANT: IKEDA, HARUO			
	APPLICANT: ISHIKAWA, JUN			
	APPLICANT: HORIKAWA, HIROSHI			
	APPLICANT: SHIBA, TADAYOSHI			
	APPLICANT: SAKAKI, YOSHIYUKI			
	APPLICANT: HATTORI, MASAHITA			
	TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES			
	FILE REFERENCE: 249-262			
	CURRENT APPLICATION NUMBER: US/10/156,761			
	CURRENT FILING DATE: 2002-05-29			
	PRIOR APPLICATION NUMBER: JP 2001-204089			
	PRIOR FILING DATE: 2001-05-30			
	PRIOR APPLICATION NUMBER: JP 2001-272697			
	PRIOR FILING DATE: 2001-08-02			
	NUMBER OF SEQ ID NOS: 15109			
	SEQ ID NO 5544			
	LENGTH: 1908			
	TYPE: DNA			
	ORGANISM: Streptomyces avermitilis			
	FEATURE:			
	NAME/KEY: CDS			
	LOCATION: (1)..(1908)			
	US-10-156-761-5544			
Alignment Scores:				
	Pred. No.:	8,31e-27	Length:	1908
	Score:	335.50	Matches:	156
	Percent Similarity:	39.12%	Conservative:	101
	Best Local Similarity:	23.74%	Mismatches:	266
	Query Match:	10.89%	Indels:	134
	DB:	15	Gaps:	26
QY	US-09-750-986D-30 (1-589) x US-10-156-761-5544 (1-1908)			
	8 LeuTrpPheProGluArgIleLeuGluAlaTrpLeuGluIleIaTrpAla-IysThrArgProGlu			


```

QY 85 PheGlyAlaMetYrAlaGlyIleProTyCysProValSerProAlaTyrSerLeuLeu 104
Db 241 CTTGCGGCCCGCATGTCGATATCGCTCGGCTGCATCACGCGGCTTATGCC---ACG 297
QY 105 SerGlnAspLeuAlaValLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
Db 298 TCGGCCCATCTCACCAACTCGCGCATCGCGGGGAGATCACCCCGGGATGTTTT 357
QY 125 AlaAlaAspAlaAlaProPheGlnArgAlaIleGlnThrIleLeuProAspAspValPro 144
Db 358 GCCAGAGATGGAGCGCCCTTTCGCGGGCGCTCGCGGAGTTTCGACGACGGTACACCG 417
QY 145 AlaIlePheThrArgGlyValLeuAlaGlyArg---ArgThrValSerPheAspSerLeu 163
Db 418 CTGGTCGGGCTCCCGCAACTCGCGAAGATGTTCCAAATACGTTTCAATTCGAAACCTCG 477
QY 164 LeuGlnGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGlyProAspThr 183
Db 478 CTTGAAACGAGCGGACCGGACCGGATGTCGCTTCATGCGGTCGCGCGCGGATACCC 537
QY 184 IleAlaIysPheLeuPheThrSerGlySerThrIysLeuProIysAlaValProThrThr 203
Db 538 GTGCGCAATTCCTTTTTCAGCTCAGGCAAGCATGCTTCGCCCAAGCGGATCATCAAGCG 597
QY 204 GlnArgMetLeuCyAlaAsnGlnIleMetLeuLeuGlnThrPheProValPheGlyGlu 223
Db 598 CAACGCATGTTGCTCCCAATCAGGAATGATGTCGATGTTATGCTACTTTCCGCGAG 657
QY 224 GlnProProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGlySerHisAsn 243
Db 658 GAGCGCGCGCTGCTGTTGATGGGCGCCAGCAACACACGCGCGGGGAGACAGGTG 717
QY 244 IleGlyIleValLeuYrAsnGlyGlyThrTyTrpLeuAspAspGlyIysProThrAla 263
Db 718 TTCAACTGCTGCTTATATGCGGTACCTATTAATTCATCGCGGCAAGCCCAAGCCG 777
QY 264 GlnGlyPheAlaGlnThrLeuArgAsnLeuSerGlnIleSerProThrAlaTyLeuThr 283
Db 778 GCGCAGATCGGCGACATCGAATCTCGGAGCAATATCGCCACTGCTATTTCCAC 837
QY 284 ValProIysGlyTrpGlnGluLeuValGlyAlaLeuGlnLysArgAspSerThrLeuArgGlu 303
Db 838 GTTCCCGCTGCGCAAGAAATGCTGTCAGGCAATGCGCAAGCGAGCGCTGTGCGCG 897
QY 304 ArgPhePheAlaArgMetIysLeuPhePhePheAlaAlaIleArgIleAspSerGlnIle 323
Db 898 AGCTTTTTCGCACTGAAAGATCTGATGATGCGGTCGCGCATGGCGCACACACA 957
QY 324 TrpAspArgLeuAspArgValAlaGlnGlnHisCysGlyGluArgIleArgMetMetAla 343
Db 958 TGGGATGCGGTGACGAGGCTTTCATGCAACGCGTGGCCATGCGTTCATCGGTGCC 1017
QY 344 GlyLeuGlyMetThrGlnThrAlaProSerCysThrPheThrGlyProLeuSerMet 363
Db 1018 GGCTCTGCTGCAAGGAGCGGCACTTCTGCTGTTCGACCGAAGCGGAGACAG 1077
QY 364 AlaGlyTyTrIleGlyLeuProAlaProGlyCysGluValIysLeuValProValAspGly 383
Db 1078 CCGGCAATTCGCTATCCGCGGCGAGGCGTGAAGTGAAGTCTCGGCTTCGATGGC 1137
QY 384 LysLeuGlnGlyArgPheHisGlyProHisValMetSerGlyTyTrpArgAlaProGlu 403
Db 1138 CCGGAGCAATTCAGGCGTGAAGCGGAGCGGACATACGCGGTTACTGGCTAATGGGAA 1197
QY 404 GlnAsnAlaGlnAlaPheAspGlyGlyTyTrpCysSerGlyAspAlaIleLysLeu 423
Db 1198 TTGACGCGCGGACGCTTCGAGGAGGAGGTTCTATGAAATCGCGCATACCGCAATTCG 1257
QY 424 AlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPhe 443
Db 1258 GCGGTGCGGAGCAATTCGCGCGGCTTTTATTCATGCGCGCATGGCGGAGAACTTC 1317

```

```

QY 444 LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
Db 1318 AAGCTGACAGACCGGACCTCGGATGCGCGTGGGCTTTCGCGGCAATTAGTCAACATG 1377
QY 464 GlyIysSerTyValLeuAspValValAlaAlaAspAspArgGlyCysLeuGlyLeu 483
Db 1378 TTCGCGGGCTGATCCCTGATGCGGTATACGCGGCAAAACCGGCGGATCGGGGCT 1437
QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyIysGluAlaSer 503
Db 1438 CTGCTGTCGATTCATACCGGCGCTACGTGAGCTTTTGCGGTTGCGCAATCTCTCC 1497
QY 504 AspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuLysArg 523
Db 1498 GATCGGAGATCATCCGCGCACCGGAGTGTGAGGCGGAGATTGCGGAAAGCTTCGCGG 1557
QY 524 LeuAsnArgGlnLysThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThr 543
Db 1558 CATCAAAACAGCGAGCGGTTCCGCAAGCGGCTCATGCGCATTCGTGATGGAAGAC 1617
QY 544 ProProSerIleAspLysGlyValThrAspLysGlySer 557
Db 1618 GC-GCTGCGCTTCGAGAGGCGAGGTTTACGGAAGAGGATCG 1658

```

```

RESULT 10
US-10-369-493-40734
; Sequence 40734, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40734
; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Caulobacter crescentus
US-10-369-493-40734

Alignment Scores:
Score: 8,89e-98 Length: 1866
Percent Similarity: 972.00 Matches: 219
Best Local Similarity: 53.32% Conservative: 86
Query Match: 38.29% Mismatches: 249
DB: 31.54% Indels: 18
Gaps: 6

US-09-750-986d-30 (1-589) x US-10-369-493-40734 (1-1866)
QY 11 ProGlyArgIleLeuGlnArgLeuGlnHisTrpAlaLysThrArgProGlnGlnThrCys 30
Db 148 CCGCGCTCGATCCCGCACCTGATGCGCGGCGGCGGCGGCGGCAATCCGCGGCTTAC 207
QY 31 ValAlaAlaArgAlaAlaAsn---GlyGlnTrpArgArgIleSerTyTrpAlaGlnMetPhe 49
Db 208 CTAGACAGCGGCAACCGAACGAGTCCCTGGCGGCGGCTCACTATGGCGAGGCGCAT 267
QY 50 HisAsnValArgAlaIleAlaGlnSerLeuLeuProTyTrpGlyLeuSerAlaGluArgPro 69
Db 268 CCGCGGCTGAGGCGCTCGCGGCGGATGATGATCAGAGCTTCAGGCCCGGACGACAGC 327
QY 70 LeuLeuIleValSerGlyAsnAspLeuGlnHisLeuGlnLeuAlaPheGlyAlaMetTyTr 89
Db 328 GTGATGATCTCTGCGGCAATTGATCGAGCAGCGCTGATGACGCTGGCGGCTTATGAC 387

```

```

QY 105 SerGlnAspLeuAlaIysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
DB 298 TCGCCGATCTCCAAACTCGCGACATCGCGGGGAGATCCCGGGGAAATGGTTT 357
QY 125 AlaAlaAspAlaIaProPheGlnArgAlaIleGluThrIleLeuProAspAspValPro 144
DB 358 GCGGAGATCGACGCCCTTTCGCCGGGCGCTCGCGCAAGTTTTCAGACGCGATCACCG 417
QY 145 AlaIlePheThrArgGlyGluLeuAlaGlyArg--ArgThrValSerPheAspSerLeu 163
DB 418 CTGGTCGGCTCCGCAACTCGCGGAAGATGTTTCCAAATGCTTTCATTTCGAAACCTCG 477
QY 164 LeuGluGlnProGlyIleGluAlaAspAlaPheAlaIaThrGlyProAspThr 183
DB 478 CTGGAACGAGCGCGACCGAAGCGTTGATGTCCTTCATCGCGTCCGCGGATACC 537
QY 184 IleAlaIysPheLeuPheThrSerGlySerThrIleLeuProIysAlaValProThrThr 203
DB 538 GTGCGCAAAATCTCTTTTCACTGACGACAGCTGTTCCGCCCAAGCGCTCATCCAGACG 597
QY 204 GlnArgMetLeuCysAlaAsnGlnMetLeuLeuGlnThrPheProValPheGlyGlu 223
DB 598 CAACGATGTTGCTCCAAATCAGAAATGATTGCGCATTTGACTCTTTCGCGAG 657
QY 224 GluProProValLeuValAspThrPleuProThrAsnIleThrPheGlyIysSerHisAsn 243
DB 658 GACCCCGCGCTGCTGTTGATGGCGCCCATGAAACACCGCGCGGGAACAGGTG 717
QY 244 IleGlyIleValIleuTyraGlnGlyIleTyrrTyrrLeuAspAspGlyIysProThrAla 263
DB 718 TTCACTCTGCTCTTATATGCGGACCTATTTATTCATCGCGCAAGCCACAGCCG 777
QY 264 GlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrrLeuThr 283
DB 778 GCGCAGATCGGCAGACACTGACAACTCCGGAACATATCGCCACCTGTTTTCAC 837
QY 284 ValProIleGlyTrpGluGluLeuValGlyAlaLeuGluArgAspSerThrLeuArgGlu 303
DB 838 GTTCCCGCTGGCCAGAAATGCTGTCACGAAATGCGCAAGCGAGCGCTGTGCCG 897
QY 304 ArgPhePheAlaArgMetIysLeuPhePheAlaIaGlyLeuSerGlnGlyIle 323
DB 898 AGCTTTTTCGCGACCTGAAGATGCTGATATGCGCTCCGCGCATCGCGACAGACA 957
QY 324 TyrAspArgLeuAspArgValAlaGluGlnHisCysGlyIleGluArgIleArgMetMetAla 343
DB 958 TGGGATGCGCTGACGAGACTTTCATGCAACGCTCGGCATGCGCTTCATGCGTGC 1017
QY 344 GlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrArgIleProLeuSerMet 363
DB 1018 GGCCTCGGCTCGACGAGACGCGACCTTCTCGCTGTTTCGACCGAACGCGACAGACA 1077
QY 364 AlaGlyTyrrIleGlyLeuProAlaProGlyCysGluValIysLeuValProValAspGly 383
DB 1078 CCGGGAATTCGGTATCCGGGCGCGAGGCGTGAAGATGAACCTCGTCCCTTCGATGCG 1137
QY 384 IysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrrTrpArgAlaProGlu 403
DB 1138 CGGTACCAATTAAGGCTGAAGGCGCGAACAATCAGCCCGGTTACTGCGCTAATGCGAA 1197
QY 404 GlnAsnAlaGlnIlaPheAspGluGluGlyTyrrTyrrCysSerGlyAspAlaIleIysLeu 423
DB 1198 TTGAAGCGGAGCTTCGACGAGAAAGGTTCTATGCAATCGCGATACCGTCAAAATTT 1257
QY 424 AlaAspProIlaAspProGluIysGlyLeuMetPheAspGlyArgIleAlaGluAspPhe 443
DB 1258 GCGCTCGGAGCATTCGCGCGCGCTTTTATTTGATGCGCGCATGCGGGAATACTTC 1317
QY 444 IysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
DB 1318 AAGCTGACAGCGGCACTGGGTGGCGCTGCGGCTTTGGCGCGCAATTAGTCMACATG 1377

```

```

QY 464 GlyIysSerTyrrValLeuAspValValAlaIaProAspArgGluCysLeuGlyLeu 483
DB 1378 TTGCGCGGCTGATCCGTATGCTCATTAACGGGGAACCGCGCGCAACTCGCGCT 1437
QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyIysGluAlaSer 503
DB 1438 CTGCTGCTGATTCATTAACCGGCGCTAGCTAGACTTGTTCGCGTTTCGCAACATCTCTCC 1497
QY 504 AspAlaGluValLeuAlaSerGluProValIleArgAlaIleThrPheAlaAspThrLeu 523
DB 1498 GATCGGAGATCTATCCGCAACCGAGCTCAGGGCGCAGATTGTCGGAAGCTTTCGCG 1557
QY 524 LeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuAspThr 543
DB 1558 CATCAAAACAGCGAGCGGCTTCGCAACCGCGTATCGCATTCCTCGATGAGAAC 1617
QY 544 ProPheSerIleAspIysGlyGluValIleAspIysGlySerIleAsnGlnArgAlaVal 563
DB 1618 GCGCTGCTGCGAAGAGCGAGGTTACGACAAAGGATCGATCAACGCGCGCGT 1677
QY 564 LeuGlnTrpArgSerAlaIysValAspAlaLeuTyrr 575
DB 1678 CTTCGATCGCAAGAGCTCGTAGATCATTTAC 1713

RESULT 9
US-10-369-493-38007
; Sequence 38007, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xiandeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 38007
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
; US-10-369-493-38007

Alignment Scores:
Score: 3,64e-126 Length: 1660
Percent Similarity: 61.19% Matches: 255
Best Local Similarity: 46.03% Mismatches: 84
Query Match: 39.78% Indels: 213
DB: 15 Gaps: 2

US-09-750-986D-30 (1-589) x US-10-369-493-38007 (1-1660)
QY 5 GluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluHisIleThrAlaIysThr 24
DB 1 GATCCGCTGGCCCTTATCGCAGAAATTAATGAACGCTCTCTTCATCTGCTGTCCTG 60
QY 25 ArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyCyluThrArgArgIleSer 44
DB 61 GCACCGAGCGCAATGAGATGCGGACCGTCAGGGTTCGAGGCAAGCATGAGCGCGGTACG 120
QY 45 TyrrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrrGlyLeu 64
DB 121 TACCGGAGGCTTCGACAAATAGCGGATGCGGCAATGCTCTTCATCATGACCTT 180
QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisIleGluIleAla 84
DB 181 TCGGTGAGAGGCGCTGCTGCTGCTTCTGAAATTCATCGAGCATGCGTTGATGCTG 240

```


QY 145 AlaIlePheThrArgGlyGluLeuAlaGlyArg---ArgThrValSerPheAspSerLeu 163
 Db 418 GTGGTGGGCTCTGCAACCTGGGAGAGATGTTCCAAATGCTTTCATTTGAAACCTTG 477
 QY 164 LeuGlnArgProGlyGlyIleGluAlaAspAsnAlaPheAlaIleThrGlyProAspThr 183
 Db 478 CTTGAAACGAGCGGACCGAGCGGTTGATCGTCTTTCATGCGGCGGCGGATACC 537
 QY 184 IleAlaIlePheLeuPheThrSerGlySerThrIleLeuProIleValIleProThrThr 203
 Db 538 GTGCGCAAAATTCCTTTTCACTGATGAGGACGATGATGCTGCGCCAAAGCGGTCATCCAGACG 597
 QY 204 GlnArgMetLeuGlyAlaIleGlnIleMetLeuGlnIleThrPheProValPheGlyGlu 223
 Db 598 CAACGATGTGGTCTCAATCAGAAATGATGCGCATTTGCGTATGCGTATCTTCCGAG 657
 QY 224 GluProProValIleuValIlePLeuPLeuProThrAsnIleThrPheGlyGlySerHisAsn 243
 Db 658 GAGCCCGGCTCTGCTGATGTTGGGCGCCATGGAACACACGCGCGGGAACAGGTG 717
 QY 244 IleGlyIleValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 263
 Db 718 TTCACCTGCTCTATATATGCGGATACCTATATATGATGCGGCAAGCCGACGCGCG 777
 QY 264 GlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaIleuLeuThr 283
 Db 778 GCGCAGATCGGCAACATGGAACATCTCGGCACTATCGCCACCTGATTTTCAC 837
 QY 284 ValProIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 303
 Db 838 GTTCCCGGCTGCGCAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 897
 QY 304 ArgPhePheAlaArgMetIleuPhePhePheAlaIleuIleuIleuIleuIleuIleu 323
 Db 898 AGCTTTTTCGCACTGAGATGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 957
 QY 324 TrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGluArgIleuMetAla 343
 Db 958 TGGGATGCGCTGCGGAGCTTTCATGCAACGCTGCGGCACTGCTGCTGCTGCTGCTG 1017
 QY 344 GlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMet 363
 Db 1018 GGGCTGCGGCTGCGAG 1077
 QY 364 AlaGlyTyrIleGlyLeuProAlaProGlyCysGluValIleuIleuIleuIleuIleu 383
 Db 1078 CCGGCAATATCGGATCCCGGCGAGGCGTGAAGATGAACTCGGCTTTCGATGGC 1137
 QY 384 LeuLeuGlyGluArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGlu 403
 Db 1138 CGGTACGAAATTCAGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
 QY 404 GlnAsnAlaGlnAlaPheAspGluGlyTyrTrpCysSerGlyAspAlaIleuLeu 423
 Db 1198 TTAAACGCGGAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
 QY 424 AlaAspProAlaAspProGlyGlyIleuMetPheAspGlyArgIleuIleuIleuIleu 443
 Db 1258 GCGCTGCGGAG 1317
 QY 444 LeuLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
 Db 1318 AACCTGAG 1377
 QY 464 GlyGlySerTyrValIleuAspValIleuAlaIleuAspArgGlyCysLeuGlyLeu 483
 Db 1378 TTGGCGGCGGCTGATCGGTATCGCTATTAACGAGAGAGAGAGAGAGAGAGAGAG 1437
 QY 484 LeuValPheProArgLeuAspCysArgAlaLeuSerGlyLeuGlyIleuValIleuSer 503
 Db 1438 CTGTGCTGCTGCTATTAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497

QY 504 AspAlaGluValIleuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuIleuArg 523
 Db 1498 GATCGGAGATCTCCGCGACCCGAGATGTCAGGCGCCAGATTCGCCGAAAGCTTTGGCG 1557
 QY 524 LeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThr 543
 Db 1558 CATCAAAACAGCGGAGCGGCTTCGCGACCGCGTCATGCGATTCCTCGATGAGAAGAC 1617
 QY 544 ProProSerIleuAspGlyGlyValIleuThrAspGlySerIleuGlnAlaVal 563
 Db 1618 GCGCTGCGCTTCAG 1677
 QY 564 LeuGlnTrpArgSerAlaIleuValIleuValIleuValIleuValIleuValIleu 575
 Db 1678 CTTCTGATCTGCAAGAGCTCTGAGATCACTTTAC 1713
 RESULT 7
 US-10-369-493-38427
 ; Sequence 38427, Application US/10369493
 ; Publication No. US2003023675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 38427
 ; LENGTH: 1728
 ; TYPE: DNA
 ; ORGANISM: Agrobacterium tumefaciens
 US-10-369-493-38427
 Alignment Scores:
 Pred. No.: 1,01e-133 Length: 1728
 Score: 1294.00 Matches: 266
 Percent Similarity: 61.19% Conservative: 84
 Best Local Similarity: 46.50% Mismatches: 220
 Query Match: 41.99% Gaps: 2
 DB: 15
 US-09-750-986d-30 (1-589) x US-10-369-493-38427 (1-1728)
 QY 5 GlnAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluHisTrpAlaIleuThr 24
 Db 1 GATCGCTTGGCCCTTATCGGAGAAATGAAATGAGACGCTTCTTCATGCTGTCGTCG 60
 QY 25 ArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTrpArgIleSer 44
 Db 61 GCACCGAGGCGACAGAGATGAGCGGACGTCAGGCTGCGAGCCATGAGCCGCGGTCAGC 120
 QY 45 TyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeu 64
 Db 121 TAGCGGAGGCTTCAGAAAGATACGCGGATCGGCAAGTCTTCTCGATCATGACCTT 180
 QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeuAla 84
 Db 181 TCGGTGAGAGGCGGCTGCTGCTGCTTTCGAAATTCATCAGCATCGATGATGATGCTG 240
 QY 85 PheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeu 104
 Db 241 CTTGCGGCGAGCATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 297
 QY 105 SerGlnAspLeuAlaIleuAspArgHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
 Db 298 TCGGCGGATCTCAACAACTCGCGGACATCCGCGGCAATCAACCCGCGGAGATGTTT 357


```

QY 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
Db 481 GCGGAACCTGCTGGGAGACCGCGGAACATCCGGAACCTGCCGCAAGCAGCAAGGATGGC 540
QY 181 ProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuProLysAlaVal 200
Db 541 CACGACCAATATCGGTCTGACGCTGTTCAACGCGGCTCGATGTATATCGACGCGCAG 780
QY 201 ProThrThrGlnArgMetLeuCyAlaAsnGlnGlnMetLeuGlnThrProVal 220
Db 601 ATCAATACGAGCGGATGATCTGCGCAATCAGTATGATCCGCGAGGAGTGGCTTC 660
QY 221 PheGlyGluGluProProValIleValAspTrpLeuProTrpAsnIleThrPheGly 240
Db 661 CTAAGAAGACGAGCGCGGATGATGTCGATCGGTCGCGGAATACACCTTCGGGCGC 720
QY 241 SerHisAsnIleGlyIleValIleuTyraAsnGlyGlyThrTyrrTyLeuAspAspGlyLys 260
Db 721 AACCAAAATATCGGTCTGACGCTGTTCAACGCGGCTCGATGTATATCGACGCGCAG 780
QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGlnIleSerProThrAla 280
Db 781 CCGACGCGGCGCGGATGCTCTCCACCATCCGCAATCGCGGATCGCGCGACGCGTG 840
QY 281 TyrLeuThrValProLysGlyTrpGluGluLeuValGlyAlaLeuGluArgAspSerThr 300
Db 841 TATTTCAAGCTTCGGAAGGCTACGACTCGCTGTCGCGGTCGTCGCGCAAGACACAG 900
QY 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaIleGlyLeuSer 320
Db 901 TTCGCAAAATGTTCTTCACGCGGCTCGATCGCATGTTCTTCGCGCGACGCTCGCG 960
QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGluIleHisCySerGlyGluArgIleArg 340
Db 961 GCGCATGTCGGAACGCGCTTGACGAGCTCGGTCGCGAGACGCGCGCGGCTGCGC 1020
QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCyThrPheThrThrGlyPro 360
Db 1021 ATGCTCACCGGCTTGCGCGCACCGAGACCGCGCTTTCATGTCGATGACCGCCAG 1080
QY 361 LeuSerMetAlaGlyTyrrIleGlyLeuProAlaProGlyCySerGlyValLysLeuValPro 380
Db 1081 ACCAGTCGCTCGGCGCATGTCGCGCTCGCGTCCGCGCAACGAGCGCAAGCTGTCGCG 1140
QY 381 ValAspGlyLysLeuGluArgPheHisGlyProHisIleValMetSerGlyTyrrTrpArg 400
Db 1141 AACAAACGCGAAGCTCGAAGTCGCGCGCAAGGCGCGAATCACCCCGGCTATTTGCGC 1200
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrrTyrrCySerGlyAspAla 420
Db 1201 GCGCGCGAGCTGACCGCATTAAGGCTTGCAGAGAGAGGCTTTCATCAAGCTCAACGATGCG 1260
QY 421 IleLysLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAla 440
Db 1261 CTGAAGCGCGGTGATGCAAGACCTTTCGCGCGCTTCATTTGCAAGCGCGGATCTCG 1320
QY 441 GlnAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 1321 GAAGACTTCAAACTGCGGTGCGGCGACCTGCGTCCGCTCGCTCAACGCGCAAGCTTC 1380
QY 461 ValLeuGluGluGlySerTyrrValIleuAspValValAlaAlaProAspArgGlyLys 480
Db 1381 ATTTGCCCGCTCGCGCTCTCGGTGCGGAGTGTGTGTCGCGGCTCGACCGCGATTTAC 1440
QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAsp-----CysArgAlaLeuSerGly 497
Db 1441 GTACACGCGGTGGG-----ATCCTGATCCGACGCGCTGCAAGCTGATCAATGCG 1491
QY 498 LeuGlyLysGluAlaSerAspAlaGluValLeuAlaSerGluProValAlaArgAlaTrpPhe 517
Db 1492 ACCGTGCGGTGGAAGACTCTCGCGGATGAGCGCGCAACATCATGATCCGCGAGGCTTTC 1551

```

```

QY 518 AlaAspTrpLeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrp 537
Db 1552 CGCGAGCGCTTCGCCACGCTGCTGAGCGAGCGCACCGGCTGCTGTCACACCGCGTACCGCGC 1611
QY 538 ValGlyLeuLeuAspThrProProSerIleAspLysGlyGluValAlaThrAspLysGlySer 557
Db 1612 GCGGTGCTGCTCGGCGAACCCTGCTGATGACAGAGGAGAGATCACCGAGAGGCTTCG 1671
QY 558 IleAsnGlnArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaLeuTyrr 575
Db 1672 GTCAACGAGCGGCGCGCTGCAATATCGCGCTCGTTGATCCGGATCTTTAC 1725

RESULT 6
US-10-369-493-35384
; Sequence 35384, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIORITY FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 35384
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-35384

Alignment Scores:
Pred. No.: 1.01e-133 Length: 1728
Score: 1294.00 Matches: 266
Percent Similarity: 61.19% Conservative: 84
Best Local Similarity: 46.50% Mismatches: 220
Query Match: 41.99% Indels: 2
DB: 15 Gaps: 2

US-09-750-986d-30 (1-589) x US-10-369-493-35384 (1-1728)
QY 5 GlnAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluHisTrpAlaLysThr 24
Db 1 GATCGCTTGCGCCCTATCCGCAAGAAATGAAATGAGGTCTCTTCACTGATGCTGTCG 60
QY 25 ArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTrpArgArgIleSer 44
Db 61 GCAACGAGGCGCAATGATGATGAGCGGACCGTCAAGGTCGCAAGCATGAGCGCGGATGAC 120
QY 45 TyrAlaGluMetPheHisAsnValArgAlaAlaAlaGlnSerLeuLeuProTyrrGlyLeu 64
Db 121 TACGCGAGGCTTCGCAAGATACGCGCGATGCGGCAATGCTTCTTCATCATGACTT 180
QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeuAla 84
Db 181 TCGGTGAGAGGCGCGCTGCTGCTTTCGAAATTCCTATCGAGTCAGTGTGATGTCG 240
QY 85 PheGlyAlaMetTyrrAlaGlyIleProTyrrCysProValSerProAlaTyrrSerLeu 104
Db 241 CTTCGCGCGCAGATGTCGATATGCGCTCGGCTGCAATCGCGCGGCTTAAGCC--ACG 297
QY 105 SerGlnAspLeuAlaLysLeuAlaArgHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
Db 298 TCGGCGCATCTTCAACAACTCGCGCATCGCGGCGAGATCACCCCGGAGATGTTT 357
QY 125 AlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspArgValPro 144
Db 358 GCCGAGATGCAAGCCCTTTCGCGGCGCTCGGCGAGGTTTTCAGACGATACACCG 417

```

QY 201 ProThrThrGlnArgMetLeuCyAlaAsnGlnGlnMetLeuGlnInThrPheProVal 220
Db 21888 CCGACTACTAGGAGGAGTCTGTGGCCAAATGCTTCTGCAAACTTCTCCGGTT 21947
QY 221 PheGlyGlnGlnProProValLeuValAspTrpLeuProTrpAsnHisThrPheGly 240
Db 21948 TTGGTGAAGAGCGCGGCTGTGTGAGTGTGCTGTGCGTGAACCACTTCCGGCGC 22007
QY 241 SerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspArgIleVal 260
Db 22008 AGCCACAACTCGCATCGTGTGTGAACAACGCGCGCATCTTGAACGACGCTAAA 22067
QY 261 ProThrAlaGlnGlnPheAlaGlnThrLeuAspAsnSerGluIleSerProThrAla 280
Db 22068 CCAACGCGCCAAAGGATTCGCGAGACCTTGGCAACTTGAAGCAAACTCTCCACCTGCG 22127
QY 281 TyrLeuThrValProGlyGlyTrpGlnGlnValGlyAlaLeuGlnArgAspSerThr 300
Db 22128 TACCTCACTGTGCGAAGAGCTGGAGAGATTAGTGGTGCCTTGAAGCAAGACAGTACC 22187
QY 301 LeuArgGlnArgPhePheAlaArgMetIleLeuPhePhePheAlaAlaIleGlyLeuSer 320
Db 22188 CTGCGCGAAGCTTCTTCTGCGCATGAAGCTTCTTCTTCCGCGCGCGCTGGGTTGTCG 22247
QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGlnGlnHisCysGlyGlyArgIleArg 340
Db 22248 CAAGGAGTCTGGAGTCTTGGACCGGGTGGCTGAACAGCATGTGGTGAAGCGATTCGC 22307
QY 341 MetMetAlaGlyLeuGlnGlyMetThrGlnThrAlaProSerCysThrPheThrThrGlyPro 360
Db 22308 ATGATGGCGGGTCTGGCGCATGAGAGATGCTCTTCTGCACTTTTACCAACGAGCCG 22367
QY 361 LeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGlyValIleValPro 380
Db 22368 CTGTCGATGCTGTGTTACATTTGGGCTGCCAGCGCTGCTGGCGAGCTCGTTCCG 22427
QY 381 ValAspGlyIleValLeuGlnGlyArgPheHisGlyProHisValMetSerGlyTyrTrpArg 400
Db 22428 GTGAGTGGAGAAATTTGAGAGGCGTTTTCATGTGTCCGACGTCATAGACGGCTACTGGCGT 22487
QY 401 AlaProGlnGlnAsnAlaGlnAlaPheAspGlnGlnGlyTyrTyrCysSerGlyAspAla 420
Db 22488 GCTCTGAACAAATGCGCAAGGCTTGCAGAGAGGCTTATCTCTCCGGGATGCC 22547
QY 421 IleValLeuAlaAspProAlaAspProGlnGlyIleValMetPheAspGlyArgIleAla 440
Db 22548 ATCAAAATTTGCGAGATCTTGGCGCATCTCAGAAAGCTGTGATGTTGACGCGTCGATTCGT 22607
QY 441 GlnAspPheIleValSerSerGlyValPheValSerValGlyProLeuArgTrpArgAla 460
Db 22608 GAAAGACTTCAAGCTGTCTCAAGGGGATTTGTTCAGCGCTGGGCCATTTGCGACGCGGGCG 22667
QY 461 ValLeuGlnGlnGlySerTyrValLeuAspValValAlaAlaProAspArgGluCys 480
Db 22668 GTTCTGGAAGCGGCTTATGCTGTCTGGAAGTGTGCTGTCTCTGATCGTGAATGC 22727
QY 481 LeuGlyIleLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyVal 500
Db 22728 CTGGATGTCTGCTGTTCGCGCTTCTCGCATGCGCGCTGTGTGTGGGGCTGAGGAAAA 22787
QY 501 GlnAlaSerAspAlaGlnValIleValAlaSerGlnProAlaArgAlaTrpPheAlaAspTrp 520
Db 22788 GAGCGCTGGACCGCGAGGCTTGGCCAGTGAAGCGGCTGGGGCTGTGTTGCTGACTGG 22847
QY 521 LeuIleValPheAlaAspArgGlnAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeu 540
Db 22848 CTCAAAGCATCATGAGAGAGCATGGCAATGCGCATCATGTGGGTAGGGCTC 22907
QY 541 LeuAspThrProProSerIleAspIleValGlyValIleThrAspGlySerIleAsnGln 560
Db 22908 CTGAGATGCGCGCTGTGATTTGATTAAGGGGAGGCTCATGCAAGGGGCTCGATTCACAG 22967

QY 561 ArgAlaValLeuGlnInThrPargSerAlaIleValAspAlaLeuTyrArgGlyLeuArgIn 580
Db 22968 CGCGCTGTTTGGAAATGCGCGTGGCGAAAGTTGATGCTGTATGTGTGAAGTCAA 23027
QY 581 SerMetLeuArgAspGlnAlaThrLeu 589
Db 23028 TCCATGCTGCTGAGCAGAGCCACACTG 23054
RESULT 5
US-10-369-493-44177
; Sequence 44177, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ. ID NOS: 47374
; SEQ. ID NO 44177
; LENGTH: 1725
; TYPE: DNA
; ORGANISM: Rhodospseudomonas palustris
US-10-369-493-44177

Alignment Scores:

Pred. No.: 6.8e-152 Length: 1725
Score: 1457.00 Matches: 292
Percent Similarity: 65.40% Conservative: 86
Best Local Similarity: 50.52% Mismatches: 194
Query Match: 47.27% Indels: 6
DB: 15 Gaps: 2

US-09-750-986D-30 (1-589) x US-10-369-493-44177 (1-1725)

QY 1 MetArgSerLeuGlnAlaLeuProPheProGlyArgIleLeuGlnArgLeuGlnHis 20
Db 1 GTCGCTCAACCAACCAACGCTCACCAGACTATCCGTCGGATTTACGACCGGCTGCATCAC 60
QY 21 TrpAlaIleThrArgProGlnGlnInThrCysValAlaAlaArgAlaAlaAsnGlyIleTrp 40
Db 61 TTCGCGAGACGCGCGCGCGCGCGGCTTTCATGCTGCGCGAGCGGAACGCTGAGGCGGCTGG 120
QY 41 ArgArgIleSerTyrAlaGlnMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 121 CGCAGATCAGCTATCCAGATGCTGCCGCGCGCCAGACCACTCCCTCGGCTGATC 180
QY 61 ProTyrGlyLeuSerAlaGlnArgProLeuLeuIleValSerGlyAsnAspLeuGlnHis 80
Db 181 GCCGCGGAGCTGCGCGCGCAACGCGCGTGAATTTCTGCGGTAATTCATGACACAT 240
QY 81 LeuGlnLeuAlaPheGlnAlaMetTyrAlaGlyIleProTyrCysArgProValSerProAla 100
Db 241 GCGATGTGATGTTCCGCGCGCTGTATGCGGGCTGCGCGATGTGCTCCGCTGCGCGCGC 300
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaIleValArgHisGlyValGlyLeuGlnPro 120
Db 301 TATTCGTGTGTTCAGAGATTACGCAAGCTGCGCATATCTCGGGCTGACTGCGC 360
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGlnThrIleLeuPro 140
Db 361 GGGCTATCTTCGCGCATGACATCACTGCTTCCGCGCGCATCTCCGCAACCGGCGC 420
QY 141 AspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyArgArgThrValSerPhe 160
Db 421 GAGATGTGAGCTCTGCGCACGCGCGGAGGTGAAGGGGCGCAAGGTGACGTGCTT 480

```
QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
Db 22068 CCAACCGCCCAAGGATTCGCGAGACGCTTCGCAACTGACGAATCTCCACCTCG 22127
QY 281 TyrLeuThrValProlylGlyTyrGluGluValGlyAlaLeuGluArgAspSerThr 300
Db 22128 TACCTCACTGCGCGAAAGGTGGAGGAATTAATGAGTCCCTTGACCGAGCACTACC 22187
QY 301 LeuArgGluArgPhePheAlaArgMetLeuPhePhePheAlaAlaAlaGlyLeuSer 320
Db 22188 CTGGCGCAACGCTTCTTCGCTCGCATGAGCTGTTCTTCTTCGCGCGGCTGGGTTGCG 22247
QY 321 GlnGlyIleTyrPheArgLeuAspArgValAlaGluGlnHisCysGlyGlyAlaArg 340
Db 22248 CAAGGAGATCGGAGATCGTTTGACCGGGTCCGTGAACAGACTGAGGTGAGCGCATTCGC 22307
QY 341 MetMetLeuAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360
Db 22308 ATGATGGCGGGTCTGGGCATGACGAGACTGCTCTTCTCTGCACTTTTACCAACCGGACCG 22367
QY 361 LeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGluValLeuValPro 380
Db 22368 CTGCGATGGCTGTTACATTGGGCTGCGACGCTGGCTGGCTGCAAGTCAAGCTGTTCCG 22427
QY 381 ValAspGlyLeuLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTyrArg 400
Db 22428 GTCGATGGAAATGGAAGGGCGTTTCATGTCGCGACGTCATGAGCGGCTTACGGGGT 22487
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAla 420
Db 22488 GCTCCCTAACAAATGCGCCAGAGCTTGCACGAGAGAGCTATTACTGCTCGCGGAGATGCC 22547
QY 421 IleValSerLeuAlaAspProAlaAspProGlnLeuGlyLeuMetPheAspGlyArgIleAla 440
Db 22548 ATCAAAATGGCGAGATCTTCGCGATCTTCAGAAAGCTCTGATGTTGACGGTCCGATTTGCT 22607
QY 441 GlnAspPheIleLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 22608 GAAAGACTTCAAGCTGTCCTCAGGGGTATTGTGCAAGCTGGGCGCATTCGCGACGGGCGC 22667
QY 461 ValLeuGluGlyGlySerTyrValLeuAspValValAlaAlaProAspArgGluCys 480
Db 22668 GTTCTGGAAGCGGCTCTTACGTCCTGACCTAGTGTGCTGCTCTGATCGTGAATGC 22727
QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
Db 22728 CTTGGATTGCTCGTTCCTCCGCTCTTCTGCACTGCGGCTTGTGCGGGCTAGGAAA 22787
QY 501 GlnAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTyrPheAlaAspTyr 520
Db 22788 GAGGCGTCGAGCGCGGAGTCTTGGCAGTGAAGCGGCTGGGCTGGTTTGGTGAATCGG 22847
QY 521 LeuValArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeu 540
Db 22848 CTCAAAAGACTCATTCAGAAAGCACTGGCAATGCGCATGCGCATCATGTGGGTAGGGCTC 22907
QY 541 LeuAspThrProProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGln 560
Db 22908 CTGCAATCGCGCGCTGATTTGATTAAGGCGAGGCTGCTGCAAGGGGCTGATTAACGAG 22967
QY 561 ArgAlaValLeuGlnTyrArgSerAlaLysValAspAlaLeuTyrArgGlyGlyAspGln 580
Db 22968 CGGCGCTTTTTCGAATGCGCGGTGGCGAAAGTTGATCGGCTGATCGTGGTGAAGATCAA 23027
QY 581 SerMetLeuArgAspGluAlaThrLeu 589
Db 23028 TCCATGCTGCGTGAAGAGCCACACTG 23054
```

RESULT 4

```
US-09-750-986D-1
; Sequence 1, Application US/09750986D
; Publication No. US20030228670A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Steinbuechel, Alexander
; APPLICANT: Priefert, Horst
; APPLICANT: Rabenhorst, Jurgen
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
; TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
; FILE REFERENCE: Bayer-9998-CAO
; CURRENT APPLICATION NUMBER: US/09/750,986D
; CURRENT FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 32679
; TYPE: DNA
; ORGANISM: Pseudomonas sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3146)..(3997)
; OTHER INFORMATION: gene = "ORF1"
US-09-750-986D-1
```

Alignment Scores:

Pred. No.:	Score:	Length:
0	3082.00	32679
Percent Similarity:	100.00%	Matches: 589
Best Local Similarity:	100.00%	Conservative: 0
Query Match:	100.00%	Mismatches: 0
DB:	10	Indels: 0
		Gaps: 0

US-09-750-986D-30 (1-589) x US-09-750-986D-1 (1-32679)

```
QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyValArgIleLeuGluArgLeuGluHis 20
Db 21288 ATGGGTTCTCTCAGAGGGCTTCTTCCCTCCCGGTCGAATTTCTTGAGCGTTCGACCAT 21347
QY 21 TrrAlaLysThrArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTyr 40
Db 21348 TGGGCTAAGACCCGTCAGAAACAAACCTGCTGCTGCTCCAGGGCGGCAATGGGGAATGG 21407
QY 41 ArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 21408 CGTGTATCAGCTACGCGGAATGTTCCACACGTCGCGCCATCGCACAGAGCTTGCTT 21467
QY 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 21468 CTTTACGCACTATCGGACAGCGCTCGCTCTTATCTCTGGAATGACTGGAAACAT 21527
QY 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAla 100
Db 21528 CTTCAAGCTGCAATTTGGGGCTATGTATGCGGCAATTCCTTATGCCCGGTGTCTCTGCT 21587
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnPro 120
Db 21588 TATTCACTGCTGTCGCAAGATTGGCGAAGCTGCTGACATCGTAGGCTTTCTGCAACCG 21647
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
Db 21648 GGACTGGTCTTCTGCGCATGACGACCTTTTCAAGCGCCCAATTGAGCACTTTCGCG 21707
QY 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPhe 160
Db 21708 GACGACCTGCGCGCAATCTTCACTCGAGGCAATTTGGCCGGGCGGCGACGCTGAGTTT 21767
QY 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
Db 21768 GACAGCTGCTGGAGGAGCGCTGGTGGGATTGAGGACAGATATATGCTTTGGGCAACTGGC 21827
QY 181 ProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuPheProlylAlaVal 200
Db 21828 CCCGATACGATTGCGCAAGTTCTTGTTCACTTGGCTCTTACCAAACTGCTTAAGGCGGTG 21887
```

Db 1321 GAAGACTTCAAGCTGTCCTCAGGGGATATTGTGACGCTTGCGCATTTGCCACGCGGGCG 1380
Qy 461 ValLeuGluGlyGlySerTyrValLeuAspValValAlaAlaProAspArgGluCys 480
Db 1381 GTTCTGGAAAGCGCGCTCTTAACGTCTGAGACGTAGTGTCTCTCTCTGATCGTGAATGC 1440
Qy 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLeu 500
Db 1441 CTTCGATTGCTGCTGTTTCCGCGCTCTTCTGACCTCGCTCTTCTGCGGGCTGAGAAA 1500
Qy 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTyrPheAlaAspTyr 520
Db 1501 GAGCGCTGGAGCCGCGAGGTGCTTGCCAGTGAAGCGGTTGCGGCTGTTGCTGACTGG 1560
Qy 521 LeuLeuArgLeuAspArgGluAlaThrGlyAspAlaSerArgGlyLeuGlyLeu 540
Db 1561 CTCAACGACTCACTCAACGAGCACTGCGCAATGCGCACTGCACTGATGAGTGGGCTC 1620
Qy 541 LeuAspThrProProSerIleAspGlyGlyValThrAspGlySerIleAspGln 560
Db 1621 CTGGATACCGCCGCGCTGATTTGATTAAGCGGAGGTCTCTACAAAGGCTGCACTACAG 1680
Qy 561 ArgAlaValLeuGlnTyrPargSerAlaValAspAlaLeuTyrArgGlyGluAspGln 580
Db 1681 CGGCGTGTGTCATAGCGCGGTGCGGCAAGTTGATGCTGTATGCTGTGAAGATCAA 1740
Qy 581 SerMetLeuArgAspGluAlaThrLeu 589
Db 1741 TCCATGCTGCTGACGAGGCGACACTG 1767

RESULT 3

US-08-976-063C-1
Sequence 1, Application US/08976063C
Publication No. US20020182697A1

GENERAL INFORMATION:

APPLICANT: Alexander Steimbuchel; Horst Priefert; Jurgen Rabenhorst
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
TITLE OF INVENTION: CONFERRAL ALCOHOL, CONFERRALDEHYDE, FERRULIC ACID, VANILLIN A
TITLE OF INVENTION: ACID AND THEIR USE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: HP VECTOR

OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,063C

FILING DATE: 21-NOV-1997
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 196 49 655.1 (Germany)

FILING DATE: 29-NOV-1996
ATTORNEY/AGENT INFORMATION:

NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141

REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
TELECOMMUNICATION INFORMATION:

TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844

TELEX:
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 32679 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (Genomic)

HYPOTHEtical: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Pseudomonas sp.

STRAIN: HR199

FEATURE:

NAME/KEY: CDS

LOCATION: 3146..3997

OTHER INFORMATION: /gene= "ORF1"

US-08-976-063C-1

US-09-750-986D-30 (1-589) x US-08-976-063C-1 (1-32679)

Alignment Scores:

Score:	Length:
3082.00	32679
Percent Similarity:	Matches:
100.00%	589
Best Local Similarity:	Conservative:
100.00%	0
Query Match:	Mismatches:
8	0
DB:	Indels:
	0
	Gaps:
	0

Qy 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluHis 20
Db 21288 ATGCGTCTCTCGAGGCGCTTCCCTTCCGCGGTGCAATTTGAGCGTCTCGAGACT 21347

Qy 21 TrpAlaIleThrArgProGluGlnThrCysValAlaAlaArgAlaIleAsnGlyGlyTyr 40
Db 21348 TGGGCTTAAGCCCGCTCAGAAACAACTCGTTGCTGCCAGGGCGGCAATGGGAAATGG 21407

Qy 41 ArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 21408 CGTCGATCAGACGACCGGAAATGTTCCAAAGTCCGGGCATGCCACAGAGCTTGCTT 21467

Qy 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 21468 CTTTACGAGCTATCGGACAGGCGCTCCGCTTATGCTCTGGAATGACCTGGAACAT 21527

Qy 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAla 100
Db 21528 CTTCAGCTGCAATTTGGGCTATGATGCGGGAATTCCTATGCCCCGGTCTCTGCT 21587

Qy 101 TyrSerLeuLeuSerGlnAspLeuAlaIleValArgHisIleValGlyLeuLeuGlnPro 120
Db 21588 TATTCACTGCTGTCGAAATTTGGGAGCTGCGTCACTGATGCTTCTGCAACCG 21647

Qy 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
Db 21648 GGAAGTGTCTTGTCTCCGATGACGACCTTTCACGCGCAATTGAGACCATTTGCGG 21707

Qy 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyValArgThrValSerPhe 160
Db 21708 GACGAGTGGCCGCAATCTTCACTCGAGGGAATTTGCGGCGGCGCAAGGAGTTT 21767

Qy 161 AspSerLeuLeuGlnProGlyIleGlyAlaAspAsnAlaPheAlaIleThrGly 180
Db 21768 GACAGCTGCTGAGAGACGCTGTTGAGGAGGATTAATCCCTTGGGCAACTGGC 21827

Qy 181 ProAspThrIleAlaIlePheLeuPheThrSerGlySerThrIleLeuProValAla 200
Db 21828 CCGATACGATTCGCAAGTCTTGTTCACCTTGGCTTACCAACTGCTTAAGGCGGTG 21887

Qy 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnIleMetLeuLeuGlnThrPheProVal 220
Db 21888 CCGACTACTAGGAAAGCTCTGCGCAATCAGCAATGCTTTCGCAAACTTCCCGGTT 21947

Qy 221 PheGlyGluGluProProValLeuValAspTyrLeuProTyrAsnHisIleThrPheGly 240
Db 21948 TTTGGTGAAGGCGCGGCTGCTGAGTGTTCGAGTTCGAGAAACCACTTCCGCGG 22007

Qy 241 SerHisAsnIleGlyIleValIleValIleValIleValIleValIleValIleValIle 260
Db 22008 AGCCACAACATCGGATCGTGTGTACAAACGCGGACGATCACTTGAAGCGGTAA 22067

QY 521 LeuIyAArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeu 540
Db 1561 CTCAAAGCATCATGAGAGACAGCACTGGCAATGCCATGCATCATGTGGATAGGGCTC 1620
QY 541 LeuAspThrProProSerIleAspIlyGlyGluValThrAspIlySerIleAsnGln 560
Db 1621 CTCGATACGCCGCCGCTCGATTGATTAAGGGCGAGGTCACTGACAAAGGCTCGATCAACACAG 1680
QY 561 ArgAlaValLeuGlnTrpArgSerAlaValAspAlaLeuTrpArgGlyIyAspGln 580
Db 1681 CGCGCTGTTTGGAAAGCCGCTCGCGCAAAAGTTGATGCGCTGATCGTGGTGAAGATCAA 1740
QY 581 SerMetLeuArgAspGluAlaThrLeu 589
Db 1741 TCATGCTGCTGACGAGGCCACACTG 1767
RESULT 2
US-09-750-986D-29
; Sequence 29, Application US/09750986D
; Publication No. US20030228670A1
; GENERAL INFORMATION:
; APPLICANT: Steinbuechel, Alexander
; APPLICANT: Pfeleert, Horst
; APPLICANT: Rabenhorst, Jurgen
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONFERYL
; TITLE OF INVENTION: ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; FILE REFERENCE: VANILLIC ACID AND THEIR USE
; CURRENT APPLICATION NUMBER: US/09/750, 986D
; PRIOR FILING DATE: 2000-12-28
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Pseudomonas sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1767)
; OTHER INFORMATION: product = "Ferulasaure-CoA-Synthetase" / gene =
; OTHER INFORMATION: "fcs"
US-09-750-986D-29
Alignment Scores:
Pred. No.: 0 Length: 1770
Score: 3082.00 Matches: 589
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
US-09-750-986D-30 (1-589) x US-09-750-986D-29 (1-1770)
QY 1 MetArgSerLeuGlnAlaLeuLeuProPheProGlyValArgIleLeuGlnArgLeuGluHis 20
Db 1 ATCCGCTTCTTCAGAGCGCTTCTTCCCTCCCGGCTGAAATCTTGAAGCTTCAGACAT 60
QY 21 TrpAlaValThrArgProGluGlnTrpCysValAlaAlaArgAlaAlaAsnGlyIyTrp 40
Db 61 TGGGCTAAGACCCGCTCCAGAAACAACCTGCTGCTCCAGGGCGGCAATGGGGAATGG 120
QY 41 ArgArgIleSerTrpAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 121 CGTCGATACGCTACGCGGAAATGTTCCACACGTCGCGCATCGCACAGACTTGCTT 180
QY 61 ProTrpGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 181 CTTTACGAGCATTCGCGAGCGCTCGCTCTATCTCTGGAATGACCTGGAACT 240
QY 81 LeuGlnLeuAlaPheGlyAlaMetTrpAlaGlyIleProTrpCysProValSerProAla 100

Db 241 CTTCAGCTGCATTTGGGGCTATGTATGCCGGCATTTCCCATATGCCCGGTCTCCGCT 300
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaIyLeuArgHisIleValGlyLeuGlnPro 120
Db 301 TATTCATGCTGTGCGAAGATTGGCGACCTCGTCACTGATCGTGGCTTCGCAACCG 360
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluTrpIleLeuPro 140
Db 361 GAGCTGGTCTTCTGCTCCGATGACGACCTTTCAGAGGCGCAATTGACCATTTCTGCCG 420
QY 141 AspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyValArgThrValSerPhe 160
Db 421 GACGAGCTGCCCAATCTTCACTCGAGGAATTGGCCGGGGCGGCACAGGTGAAGTTT 480
QY 161 AspSerLeuLeuGlnProGlyIyIleGluAlaAspAsnAlaPheAlaIleThrGly 180
Db 481 GACAGCTGCTGAGAGACCTGCTGGGATTTGAGGCAGATATATGCTTTGGCGCACTGGC 540
QY 181 ProAspThrIleAlaIyAspPheLeuPheThrSerGlySerThrIyAspProIyAlaVal 200
Db 541 CCGGATACGATTGCCAAGTTCTTGTCACTTGGCTCTACCAACCTGCTTAAGGGGCTG 600
QY 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnIleMetLeuLeuGlnTrpPheProVal 220
Db 601 CCGACTACTCAGCGGAATGCTCTGCGCAATCACAGATGCTTCTGCAAACTTCCGGTT 660
QY 221 PheGlyGlnGluProProValIleuValAspTrpLeuProThrAsnIleThrPheGly 240
Db 661 TTGGTGAAGAGCGCGGCTGCTGTGAGCTGTGCTGTGCAACACACTTCCGGGCGC 720
QY 241 SerHisAsnIleGlyIleValLeuTrpAsnGlyIyThrTrpTrpLeuAspAspGlyIy 260
Db 721 AGCCACAAATCGCGCATCGTGTGTAACAAGCGGCGACGATACCTTGACGACGGTAA 780
QY 261 ProThrAlaGlnGlyPheAlaGluTrpLeuArgAsnLeuSerGluIleSerProThrAla 280
Db 781 CCAACCGCCCMAGGTTCCGCGAGACGCTTCGCAACTTGAGCCAAATCTCTCCACATGCG 840
QY 281 TyrLeuThrValProIyAspGlyTrpGluIleuValGlyAlaLeuGlnArgAspSerThr 300
Db 841 TACTCACTGTGCGAAGAGCTGGAGGAATTGAGGTGCTTGTAGCCAGACATGACC 900
QY 301 LeuArgIyAspPhePheAlaArgMetIyLeuPhePhePheAlaAlaIyLeuSer 320
Db 901 CTGCGGAACGCTTCTTCGCTCGCATGAAGCTGTCTTCTTCGCGGCTGGCTGTGCTG 960
QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGlnGlnHisCysGlyGlyIyAlaArg 340
Db 961 CAAGGATCTGGGATGTTTGGACCGGGTCCGCTGAACAGCATGTGTGAGCCCATTCGC 1020
QY 341 MetMetAlaGlyLeuGlnIyMetThrGluTrpAlaProSerCysThrPheThrTrpIyPro 360
Db 1021 ATGATGCGGGGTGTGGACATGACGAGACTGCTCTTCTCTGCACTTTTACACCGAGACG 1080
QY 361 LeuSerMetAlaGlyTrpIleGlyLeuProAlaProGlyCysGluValIyLeuValPro 380
Db 1081 CTGTGATGGCTGTGATACATTTGGCTGCGACGCGCTGGCGCAAGGTCAGTTCGCG 1140
QY 381 ValAspGlyIyLeuGlnGluIyArgPheHisGlyProHisIyValMetSerGlyTrpArg 400
Db 1141 GTCGATGGGAAATTTGAAAGGCGTTTCATGTGCTCCACGCTCATGAGCGGCTCTGGCGT 1200
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTrpTrpCysSerGlyAspAla 420
Db 1201 GCTCTGAACAAATATCCCAAGCTTTCAGAGGAGGCTATTATGCTCCGGGTATGCC 1260
QY 421 IleIyLeuAlaAspProAlaAspProGlnIyGlyLeuMetCysPheAspGlyArgIleAla 440
Db 1261 ATCAAAATTTGGCAGATCTCTGCGATCTTCAGAAAGCTGATGTTCACGGTGAATTCGT 1320
QY 441 GluAspPheIyLeuSerSerGlyIyAlaPheValSerValGlyProLeuArgThrArgAla 460

SOFTWARE: wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: us/09/976, 063C
FILING DATE: 21-NOV-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 196 49 655.1 (Germany)
FILING DATE: 29-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Bliccoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
TELEX:
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 1770 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1767
OTHER INFORMATION: /product=
OTHER INFORMATION: "Ferulaseure-CoA-Synthetase"
OTHER INFORMATION: /gene="fcs"
US-08-976-063C-29

Alignment Scores:
Pred. No.: 0 Length: 1770
Score: 3082.00 Matches: 589
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-09-750-986D-30 (1-589) x US-08-976-063C-29 (1-1770)

QY 1 MetArgSerLeuGluAlaLeuProPheProGlyValGlyLeuGluArgLeuGluHis 20
DB 1 ATGGGTCTCTCGAGGCGCTTCTCCCTCCGGGTAATCTTGAGCGCTCGACAT 60
QY 21 TTPAlaYerThrArgProGluGlnThrCysValAlaAlaArgAlaAlaAngLYGluTrp 40
DB 61 TGGGCTAAGACCCGTCAGAAACAAACCTGCTGCTGCCAGGCGGCAATGGGGAATGG 120
QY 41 ArgArgGlySerTYrAlaGluMetPheHisValAlaArgAlaAlaAngLYGluTrp 60
DB 121 CGTCGTATCAGCTACCGGAAATGTTCCACAACGTCGCCGCATCGCACAGAGCTTCTT 180
QY 61 ProTYrGlyLeuSerAlaGluArgProLeuLeuLeuValSerGlyAsnAspLeuGluHis 80
DB 181 CTTTACGAGCTATCGGACAGCGCTCCGCTTATCTCTTGAAATGACCTGAAACAT 240
QY 81 LeuGlnLeuAlaPheGlyAlaMetTYrAlaGlyLeuProTYrCysProValSerProAla 100
DB 241 CTTTACGCTGCAATTTGGGCTATGTATCGGGGCAATCCCTATTTGCCGGGTCTCCGCT 300
QY 101 TYrSerLeuLeuSerGlnAspLeuAlaLYsLeuArgHisAlaValGlyLeuLeuGlnPro 120
DB 301 TATTCATGCTGTCGCAAGATTGGCGAAGCTGCTGACATCGTAGGTCCTTCTGCAACCG 360
QY 121 GlyLeuValAlaPheAlaAlaAspAlaAlaProPheGlnArgAlaAlaGluTrpLeuPro 140
DB 361 GGACTGCTCTTGTGCGGATGACGACCTTCCAGGCGGCAATTGAGACCATTTCTGCG 420
QY 141 AspAspValProAlaAlaPheThrArgGlyGluLeuAlaGlyArgArgTrpValSerPhe 160

DB 421 GACGACGTCCCGCAATCTTCACTCGAGCGAATTGGCGGGCGCGCACGCTGACTTT 480
QY 161 AspSerLeuLeuGluGlnProGlyGlyLeuAlaAspAsnAlaPheAlaAlaTrpGly 180
DB 481 GACAGCTGCTGGAGCAGCTGGTGGATTGAGCAGATTAATGCTTTGGCGCACTGGC 540
QY 181 ProAspThrLeuAlaLYsPheLeuPheThrSerGlySerThrLYsLeuProLYsAlaVal 200
DB 541 CCGATACGATTGCCAAGTTCTTTGTTCACTTGCTGCTTACCAAACTGCTTAAGCGGTG 600
QY 201 ProThrThrGlnArgMetLeuCysAlaAngGlnMetLeuLeuGlnThrPheProVal 220
DB 601 CCACACTACTCAGGAAATGCTGCGCAATCAACATCTTCTGAAACTTCCCGGTT 660
QY 221 PheGlyGluGluProProValLeuValAspTrpLeuProTrpAsnHisThrPheGly 240
DB 661 TTGGTGAAGAGCGCGGCTGCTGAGCTGCTTCCCTGAAACACACTTTCGCGGCG 720
QY 241 SerHisAsnIleGlyIleValLeuTYrAsnGlyTYrThrTYrLeuAspAspGlyLYs 260
DB 721 AGCCACACATGGGATCGGTGTACACAGGCGCACGACTACCTTGACGCGGTAA 780
QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
DB 781 CCAAACCGCCCAAGGGTTCGCGAGACGCTTCGCACTTAAGCGAAATCTTCCACTGCG 840
QY 281 TYrLeuThrValProLYsGlyTrpGluGluLeuValGlyAlaLeuGlnArgAspSerThr 300
DB 841 TACCTCACTGTGCGGAAGCTGGAGAAATTATGTGGTGCCCTTACGAGACAGTACC 900
QY 301 LeuArgGluArgPhePheAlaArgMetLYsLeuPhePheAlaAlaAlaGlyLeuSer 320
DB 901 CTGCGGAAACGCTTCTTCGCTGCATGAAGCTGTTCTTCTGCGCGCGGTGGTGTG 960
QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGlyAlaGlyLeuArg 340
DB 961 CAAGGATCTGGGATGTGTGACCGGGTCGCGCAACACACTGTGTGACCATTCG 1020
QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrArgLYsPro 360
DB 1021 ATATGCGGGGTGGAGCATGACGAGACTGCTCTTCTGCACTTTTACACCGGACCG 1080
QY 361 LeuSerMetAlaGlyTYrIleGlyLeuProAlaProGlyCysGlyValLYsLeuValPro 380
DB 1081 CTGTGATGGCTGTTACATTTGGGCTGCCAGCGCTGGCGGAGGTCAAGCTCGTTCCG 1140
QY 381 ValAspGlyLYsLeuGluGlyArgPheHisGlyProHisValMetSerGlyTYrTrpArg 400
DB 1141 GTCGATGGGAATTTGAAAGGCGCTTTCATGTGCTCCGACATCAATGAGCGGCTACTGGCGT 1200
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTYrTYrCysSerGlyValAspAla 420
DB 1201 GCTCTGAAACAAATGCCCAAGGCTTCGACGAGAAAGGCTATTAATGCTCCGGATGCC 1260
QY 421 IleLYsLeuAlaAspProAlaAspProGlnLYsGlyLeuMetPheAspGlyArgIleAla 440
DB 1261 ATCAATTTGGCAATCTCTGCGATCTCAAGAAAGCTGATGTTTGAACGCTCAATTGCT 1320
QY 441 GluAspPheLYsLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
DB 1321 GAAGACTTCMAAGCTGCTCAAGGGGTATTTGTACAGGTTGGGCGCATTTGCCACGCGGCG 1380
QY 461 ValLeuGluGlyGlySerTYrValLeuAspValValAlaAlaProAspArgLYsCys 480
DB 1381 GTTCTGAAAGCGGCTCTTAAGCTTGAGAGTATGTGCTCTCTGATCTGTAATGC 1440
QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLYs 500
DB 1441 CTTGGAATGCTGCTGTTTCCGCGCTTCTGATGCGCGGTCTTGCGGGGCTAGGAAA 1500
QY 501 GluAlaSerAspAlaGluValAlaLeuAspGluProValArgAlaTrpPheAlaAspTrp 520
DB 1501 GAGGCGTCGAGCCGAGGCTTGGCAGTGAACCGGCTTCGGGCTTGTTGTCTGACTGG 1560

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 8, 2005, 02:03:17 ; Search time 791 Seconds
(without alignments)
4278.544 Million cell updates/sec

Title: US-09-750-986D-30
Sequence: 1 MRSLEALPPGRILRLER.....KVDALYRGDSQMLRDEATL 589

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 6.0 , Ygapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4300275 seqs, 2872944193 residues
Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=x1d
-O/cgnt2_1/USPTO.epool_P/US9750986/runat_03022005_071745_18678/app_query.fasta.1.775
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rmdb -MISMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bitsum62
-TRANS=human40.ccd -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09750986 @CGN 1.1 723 @runat 03022005_071745_18678
-NCPU=6 -ICPU=3 -NO_MMAR -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -18678
-LONGLOG -DEV_TIMEOUT=120 -MARR_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgnt2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgnt2_6/prodata/2/pubpna/PCRT_NEW_PUB.seq:*
3: /cgnt2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgnt2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgnt2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgnt2_6/prodata/2/pubpna/PCRTUS_PUBCOMB.seq:*
7: /cgnt2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgnt2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgnt2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgnt2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgnt2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgnt2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgnt2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgnt2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgnt2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgnt2_6/prodata/2/pubpna/US10E_PUBCOMB.seq:*
17: /cgnt2_6/prodata/2/pubpna/US10F_PUBCOMB.seq:*
18: /cgnt2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
19: /cgnt2_6/prodata/2/pubpna/US11_NEW_PUB.seq:*
20: /cgnt2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
21: /cgnt2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3082	100.0	1770	8	US-08-976-063C-29	Sequence 29, Appl
2	3082	100.0	1770	8	US-09-750-986D-29	Sequence 29, Appl
3	3082	100.0	32679	8	US-08-976-063C-1	Sequence 1, Appl
4	3082	100.0	32679	10	US-09-750-986D-1	Sequence 1, Appl
5	1457	47.3	1725	15	US-10-369-493-44177	Sequence 44177, A
6	1294	42.0	1728	15	US-10-369-493-35384	Sequence 35384, A
7	1294	42.0	1728	15	US-10-369-493-38427	Sequence 38427, A
8	1294	42.0	1728	15	US-10-369-493-38903	Sequence 38903, A
9	1226	39.8	1660	15	US-10-369-493-38007	Sequence 38007, A
10	972	31.5	1866	15	US-10-369-493-40734	Sequence 40734, A
11	383.5	12.4	1632	15	US-10-369-493-23931	Sequence 23931, A
12	335.5	10.9	1908	15	US-10-156-761-5544	Sequence 5544, Ap
13	335.5	10.9	9025608	15	US-10-156-761-1	Sequence 1, Appl
14	324	10.5	1746	15	US-10-369-493-34073	Sequence 34073, A
15	314.5	10.2	225646	17	US-10-470-565-1	Sequence 1, Appl
16	286.5	9.3	1666	15	US-10-369-493-37423	Sequence 37423, A
17	284	9.2	2103	14	US-10-119-136-124	Sequence 124, App
18	283	9.2	1830121	14	US-10-410-031-124	Sequence 124, App
19	283	9.2	1830121	14	US-10-329-960-1	Sequence 1, Appl
20	283	9.2	1830121	16	US-10-329-670-1	Sequence 1, Appl
21	283	9.2	1830121	18	US-10-158-865-1	Sequence 1, Appl
22	281	9.1	1755	15	US-10-369-493-23998	Sequence 23998, A
23	281	9.1	1839	15	US-10-369-493-43422	Sequence 43422, A
24	279	9.1	2103	10	US-09-906-419-6	Sequence 6, Appl
25	279	9.1	2103	14	US-10-119-136-6	Sequence 6, Appl
26	279	9.1	2103	16	US-10-410-031-6	Sequence 6, Appl
27	278.5	9.0	1839	15	US-10-369-493-44573	Sequence 44573, A
28	278	9.0	2106	10	US-09-906-419-5	Sequence 5, Appl
29	277.5	9.0	1515	15	US-10-369-493-35776	Sequence 35776, A
30	277.5	9.0	1794	15	US-10-369-493-31796	Sequence 31796, A
31	277	9.0	2097	9	US-09-938-8428-2269	Sequence 2269, Ap
32	277	9.0	2097	11	US-09-938-8428-2269	Sequence 2269, Ap
33	275	8.9	2257	14	US-10-119-136-5	Sequence 5, Appl
34	275	8.9	2257	16	US-10-410-031-5	Sequence 5, Appl
35	273	8.9	2287	18	US-10-739-930-1310	Sequence 1310, Ap
36	272	8.8	1767	15	US-10-369-493-44932	Sequence 44932, A
37	268	8.7	1467	15	US-10-369-493-44734	Sequence 44734, A
38	268	8.7	1500	15	US-10-369-493-43198	Sequence 43198, A
39	265.5	8.6	85632	16	US-10-461-194-1	Sequence 1, Appl
40	265	8.6	1806	15	US-10-369-493-24610	Sequence 24610, A
41	265	8.6	1914	15	US-10-156-761-4802	Sequence 4802, Ap
42	264.5	8.6	1639	10	US-09-838-469-10	Sequence 10, Appl
43	264.5	8.6	2542	18	US-10-425-115-115985	Sequence 115985,
44	264.5	8.6	9025608	15	US-10-156-761-1	Sequence 1, Appl
45	263.5	8.5	1639	10	US-09-813-279B-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-976-063C-29
; Sequence 29, Application US/08976063C
; Publication No. US20020182697A1
GENERAL INFORMATION:
; APPLICANT: Alexander Steinhuechel; Horst Prietert; Jurgen Rabenhorst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
; TITLE OF INVENTION: CONFERYL ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN AN
; NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SPRING KRAMER SCHAEFER & BRISCOE
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: HP VECTOR
; OPERATING SYSTEM: DOS

This Page Blank (uspto)

```

; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 42
; LENGTH: 1639
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant Luciferase
; US-09-396-154-42

```

```

Alignment Scores:
Pred. No.:      1,086-18      Length:      1639
Score:          258.00        Matches:      127
Percent Similarity: 36.52%    Conservative: 79
Best Local Similarity: 22.52% Mismatches:     218
Query Match:    8.37%        Indels:       141
                        Gaps:        23

```

US-09-750-986D-30 (1-589) x US-09-396-154-42 (1-1639)

```

Qy      30  CysValAlaAlaArgAlaAlaAsnGlyLutRpaRgaRglSerTyraAlaGluMetPhe 49
      119  TGCATACATTCGACCAATGCTCATACAAA---GAAATGTTTATATGAGAGATTCTG 175
Qy      50  HisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrglyLeuSerAlaGluArgPro 69
      176  AAACGTGCTGCTGCTTACGGGAAAGTTTAAAAAGTATGATTTAAAAACAAACGACACA 235
Qy      70  LeuLeuIleValSerGlyAsnAspLeuGluHisIleuGlnLeuAlaPheGlyAlaMetTyr 89
      236  ATACCGGTGTGTACCGAAATAGCTGTCAATTTCTCTCTGTAATTCATCTATTGTAT 295
Qy      90  AlaGlyIleProTyrcysProvalSerProAlaTySerLeuLeuSerGlnAspLeuAla 109
      296  CTTCGATTAATTTGTGGCCCTGTACGATTAATATC-----ATTGAACGTGAATTATA 349
Qy      110  LysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAspAlaAla 129
      350  -----CACAGTCTTGTGATTGTAAACACGACATAGTTTCTCCACAGAAATACT 400
Qy      130  ProPheGlnArgAlaIleGluThrIleLeuProAspAlaProAlaIlePheThrArg 149
      401  ----TTCAAAAAGTACTGAAT-----GTAAA 424
Qy      150  GluGluLeuAlaGlyArgArgThrValSerPheAspSerLeuLeuGlnProGlyGly 169
      425  TCTAAATTAAATCTATTGAACATATTATATTAGACTTAATATGATGACTTGAAGGT 484
Qy      170  IleGluAlaAspAsnAlaPheAlaIleThrGlyProAspThr----- 183
      485  TATCAATGCGCTCAACAACATTATTCTCAAAATTCACATAGTAATCTGACGTAAAAA 544
Qy      184  -----IleAlaLysPheLeuPhePheSerGly 192
      545  TTTAAACCAATATCTTTTATCGACAGATCAGGTGGTGGTGAATTTCTTCGTGCT 604
Qy      193  SerThrLysLeuProLysAlaValProThrThrGlnArgMetLeuCybAlaAsnGln 212
      605  ACAACTGTGTGCGGAAGGAGTCAATGCTAATCAACAAGATATATTGGCA-----CGA 658
Qy      213  MetLeuLeuGlnThrPheProValPheGly-----GluGluProProValLeuVal 229
      659  TTTTCATTCGAAAAAGATCTTACTTTTGGTAAACCAATTAATCCACGTCAAGCAATTTTA 718
Qy      230  AspTrpLeuProTrpAsnHisThrPheGlyGlySerHisAsnIleGly----- 245
      719  ACGGTAAATCCTTCCACCATGATGTTGGTATGATGACCACTTAGAATACTTACTTGT 778
Qy      246  -----IleValLeuTyraAsnGlyGlyThrTyrrTyrrLeuAspAspGlyLysProThr 262

```

```

Db      779  GGATTCGAGCTTGTCTTAATGAC----- 802
Qy      263  AlaGlnGlyPheAlaGluThr-----LeuArgAsnLeuSerGluIleSerProThrAla 280
      803  -----ACGTTTGAAGAAAACTATTCTTACATTCATTAACAATTAATTAAGCGAAAGT 856
Qy      281  TyrLeuThrValProLysGlyTyrGluGluLeuValGlyAlaLeuGluArgAspSerThr 300
      857  ACTTACTGTATCA-----ACATTATGCGCATTTCTTGCAAAAAGTCACTTA 904
Qy      301  LeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaAlaGlyLeuSer 320
      905  GTTGAAGAGTACGATTTATGCACTTAAAGAAATTCATCTGTGGCGCACTTATCA 964
Qy      321  GlnGlyIleThrPaspArgLeuAspArgValAlaGluGlnHisCysGlyGluArgIleArg 340
      965  AAAGAAAT-----MetMetAlaGlyLeuGlyMetThrGluThrAlaProSer 353
Qy      341  -----MetMetAlaGlyLeuGlyMetThrGluThrAlaProSer 353
      989  AAACGTTAAATTAATTAACCTTGTACGACAGGATGATTAACAGAAACCACTTCGCT 1048
Qy      354  CysThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGly 373
      1049  GTTTTAATTCACCGAAGCGTGAACCGCAACCGGATCAACGTGTAATTAATGATCACTT 1108
Qy      374  CysGluValLysLeuVal---ProValAspGlyLysLeuGluGly----- 387
      1109  CAGCGTGTAAAGTTGTCATCTTACACAGGAAAAATTTGGGCGCAATGAACCTGCA 1168
Qy      388  -----ArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGluGlnAsn 405
      1169  GAATGTATTTTAAAGCCCGCATGATATGAGGTTATTAATTAATGAAGAACTACT 1228
Qy      406  AlaGlnAlaPheAspGluGlyTyrTyrcysSerGlyAspAlaIleLysLeuAlaAsp 425
      1229  AAGCAATTAATGATATATGACGATGCTGCTGTGAT---ATTGCTTATTATGAC 1285
Qy      426  ProAlaAspProGlnLysGlyLeuMetPheAspGlyValGlyIleAlaGluAspPheLysLeu 445
      1286  AAT-----GATGGCCATTTTATATTTGTCGACAGGCTG 1318
Qy      446  SerSerGlyVal-----PheValSerValGlyProLeuArgThrArgAlaValLeu 462
      1319  AAGTCACTGATTAATTAATTAAGTTTACAGTTGCACCTCTGAAATTTAGCGGAATCTC 1378
Qy      463  GluGlyGlySerTyrrValLeuAspValValAlaAlaProAspArgGluCybLeuGly 482
      1379  TTACAAACATCCGTATATATTGTTGATGCCGCGCTTACTGTATACCGATGAAGCCGCGGC 1438
Qy      483  LeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGluAla 502
      1439  -----GAGCTTCC-AGCTGCAAGGTGTGTGT----- 1464
Qy      503  SerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp----- 520
      1465  -----ACAGACTGGAAATA 1479
Qy      521  LeuLysArgLeuAsnArg-----GluAlaThrGlyAsnAlaSerArgIle 535
      1480  TCTAAACGAACAAATCTTACAGATTAATTTGGCCAGTCAAGTTTCAACAGCAAAATGGCT 1539
Qy      536  MetTrpValGly 539
      1540  ACGTGTGGGGGT 1551
Db

```

Search completed: February 8, 2005, 04:55:18
 Job time : 7107 secs

```

Qy      2 ArgSerLeuGluAlaLeu-----ProheProGlyArgIleLeu 15
Db      2448137 AGGTCAAGTAGAGTGTGATCGCTGAGATTAGCCTCCCGCCATTTCATCTGCGC 2448136
Qy      16 GluArg-----LeuGluHisTrpAlaYsrTrpArgProGluGln 28
Db      2448197 GAGCAGCAGCAACGTCGGCGCATGTGTGAGCAT---GACGCTGACGATCCCGACTAC 2448253
Qy      29 ThrCysValAlaAlaArgAlaAlaAsnGlyGluTrpArgArgIleSerTrpAlaGluMet 48
Db      2448254 GTC---ATCTATCAACCGCTGATGACGCGGTCTGACCGCATGTCAGCTGTGGCGAGCA 2448310
Qy      49 PheHisAsnValArgAlaIleAlaGlnSerLeuLeuProGlyLeuSerAlaGluArg 68
Db      2448311 GCCAACCAAGATGTGTGGCGGCTCTGCTGTGATTTCACTGGGGGGTGCAGCGCGCAT 2448370
Qy      69 ProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeuAlaPheGlyAlaMet 88
Db      2448371 CGGGTAGTCACTTCTCTGCGACCGCGTACGAGTGGCGCATCTCGATTTCGGGATTCTG 2448430
Qy      89 TrpAlaGlyIleProGlyCysProValSerProAlaTrpSerLeuLeuSerGlnAspLeu 108
Db      2448431 GCTGTGGGT-----GCGGTCAACCGTACCGCATCGACGACTCTGTCAGCGGAG--- 2448478
Qy      109 AlaLeuLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAspAla 128
Db      2448479 ---CAGGTGGCTGGGTTTTCACAGACTCCGAGCGGTGTGTGTGGCGAAACCGAC 2448535
Qy      129 AlaProPheGlnArgAlaIleGluTrpIleLeuProAspAspAlaProAlaIlePheThr 148
Db      2448536 TCACAGCGCAACATGCTCGCGCA---CTCTCGCGCAAGCGTGGCGCTGCGGAG 2448589
Qy      149 ArgGlyGlyLeuAlaGlyArgArgThrValSerPheAspSerLeuLeuGlnProGly 168
Db      2448590 GTACTGAGAGTCCGCGGTTCGGGTCCCAACGCTCATCGGTCTGACGAGGGCGGCGCC 2448649
Qy      169 GlyIleGluAlaAspAsn-----AlaPheAlaIleThrGlyProAsp 182
Db      2448650 TCGGTGACCGCGCGAGCTAACCGCGCGCTCGCCGCTACGCTGACGAGACCGCGC 2448709
Qy      183 ThrIleAlaLeuYsrPheLeuPheThrSerGlySerThrIleYsrProGlyAlaValProThr 202
Db      2448710 ACGCTT-----ATCTACACTCGGGACACACCGAGACCGCAAGGCTGCGCACTG 2448760
Qy      203 ThrGlnArgMetLeuCysAlaAsnGlnGlnMetLeuGlnThrPheProValPheGly 222
Db      2448761 ACCCAATCCAACTGTTTCAAGATTAAAGCGCGCATATTCACCGCGCTGCTG 2448820
Qy      223 GluGluProProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGlySerHis 242
Db      2448821 CGCAAGGCTAGCGCGCTGCTGCTTCTCTGCGCTAGCTATGTGCTG-----GCCGC 2448874
Qy      243 AsnIleGlyIleValLeuTrpAsnGlyGlyThrTrpTrpLeuAspArgIleYsrProThr 262
Db      2448875 GCATATCATGTGGCGCGCTTCCAC-----TCCAAAGTCAACC 2448910
Qy      263 AlaGlnGlyPheAlaGluTrpLeuArgAsn-----LeuSerGluIleSerPro 278
Db      2448911 GTG---GGATTCAACGACGACGATCAAGATCTGCTGCGATGTGGGGGTTCACACCG 2448967
Qy      279 ThrAlaTrpLeuThrValProGlyGlyTrpGluGluLeuValGlyAlaLeuGluArgAsp 298
Db      2448968 ACGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2449027
Qy      299 SerTrpLeuArgGluArgPhePheAlaArgMetLeuPhePheAlaAla----- 316
Db      2449028 GCCGCCAACGCC-----GCCAAAGGGGAAATCTTCGGGATCGCGCGCAGAC 2449075
Qy      317 -----AlaGlyLeu-----Ser 320
Db      2449076 GCGGTGACTGAGCGAAGCTTGGACCGCGCGGACCGGGGCTGACTGCGCGCCAG 2449135
Qy      321 GlnGlyIleTrpAspArgLeu----- 327

```

```

Db      2449136 CACCGGTGTTCACCGCGCTGTGCTTACCGCAACTGCTGCGGCTGCGGCAACTGC 2449195
Qy      328 -----AspArgValAlaGluGlnHisCysGly 336
Db      2449196 CGCGCGCGGTCTCCGCGCGCGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2449255
Qy      337 GluArgIleArgMetLeuAlaGlyLeuGlyMetThrGluTrpAlaProSerCysTrpPhe 356
Db      2449256 GCCGTCTCACCACTTACGAGGATACGCGGTGAGCGGACGATGGGGCGGTGCCATC 2449315
Qy      357 ThrTrpGlyProLeuSerMetAlaGlyTrpIleGlyLeuProAlaProGlyCysGluVal 376
Db      2449316 AGCCAGTCAATATCTTAAAGATCGGAATCTGTGGAAGCGGTGCGCGCAACACTCTA 2449375
Qy      377 LysLeuValProValAlaPheGlyIleLeuGluGlyArgPheHisGlyProHisValMetSer 396
Db      2449376 CGCATCGCGAC---GATGGCGAGCTGTGTGTGCGC-----GTTGGCGGTATTTCAGC 2449426
Qy      397 GlyTrpTrpArgAlaProGluGlnAlaAlaGlnAlaPheAspGluGluGlyTrpTrpCys 416
Db      2449427 GGTCTGTGGCGCAACGAGCTTACCGAGGCAATTC---ACCAACGCTGTGTTCAG 2449483
Qy      417 SerGlyAspAlaIleLysLeuAlaAspProAlaAspProGluIleGlyLeuMetCysAsp 436
Db      2449484 ACGGTGAT-----CTCGGTGCGGTGAGCAAGACCGGTCTTGACGATCAC 2449531
Qy      437 GlyArgIleAlaGluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeu 456
Db      2449532 GCGCGCAAGAAATTTATTCGTCACCGCGGCGGT-----AAATATGCGCGCGCT 2449585
Qy      457 ArgThrArgAlaValLeuGluGlyGlySerTrpValLeuAspValValAlaAlaPro 476
Db      2449586 GTCTGAGAACCAAGCTGGGGGCCCAACCACTGATACGCCAGCGGTGTGTGGGAC 2449645
Qy      477 AspArgGlyCysLeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSer 496
Db      2449646 GCCAAGCCCTTCACTGCGCGTTATC-----ACCATGACCTTGAAGGATTTCAG 2449696
Qy      497 Gly-----LeuGlyIleGluAlaSerAsp 504
Db      2449697 GCGTGAAGCAACGCAACAGACAGACGCTGCGCTGCGGTATTTGGCCACCGAC 2449756
Qy      505 AlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuYsrAspLeu 524
Db      2449757 CCGGATCTGATTTGCCAGATCGACGCGGCC-----GTCAAAACAGGCC 2449798
Qy      525 AsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspTrpPro 544
Db      2449799 AATCTT---CGGTGTCAATGCGCGAGTGTGATCGCAAGTTCGGAATCTGCGCGTGCAC 2449855
Qy      545 ProSerIleAspIleGlyGluValThrAspIleGlySerIleAsnGlnArgAlaValLeu 564
Db      2449856 TTCACCGAGACACCGCGCGAGTACCGCGCAATGAAGTCAACGCAAGGTGTGGCC 2449915
Qy      565 GlnTrpArgSerAlaLysValaAspAlaLeuTrpArgGlyGlu 578
Db      2449916 GAGAAGTTCCTTCGATATCGAGATTCAGATTCACACAGGAA 2449957

RESULT 15
US-09-396-154-42
; Sequence 42. Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; FILE REFERENCE: 341.012US1
; CURRENT APPLICATION NUMBER: US/09/396,154
; EARLIER FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: US 09/156,946
; EARLIER FILING DATE: 1998-09-18

```

QY 89 TyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeuSerGlnAspLeu 108
Db 2445732 GCTGTGGGT-----GGGTCACCGGTACCGACTACGAGACCTGTCAGCGAG--- 2445779
QY 109 AlaIleuLeuArgHisIleValGlyLeuLeuGlnProClyLeuValPheAlaIleAspAla 128
Db 2445780 ---CAGGTGGCTGGTGGTTTACAGAGCTCCGAGCGGTGTGTTTCCGGAACCGAC 2445836
QY 129 AlaProheGlnArgAlaIleGluThrIleLeuProAspAspValProAlaIlePheThr 148
Db 2445837 TCACACCCGACATAGTCCGCCAA-----CTCCCGGACACGCGGCCCTCCGCGGAG 2445890
QY 149 ArgGlyGluLeuAlaGlyArgThrValSerPheAspSerLeuLeuGlnProGly 168
Db 2445891 GTACTGACATGCGCGGTGGGTCACAGCGCTGCATCGGCTCAGCGAGGCGGCGCC 2445950
QY 169 GlyIleGluAlaAspAn-----AlaPheAlaAlaThrGlyProAsp 182
Db 2445951 TCGGTCACCCCGGCGGACTAACCGCCGCTCCGCCCTACGATCGACGAGCCCGCG 2446010
QY 183 ThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuProLysAlaValProThr 202
Db 2446011 ACGCTT-----ATCTACACTCGGGACACACCGGACGACCCAGGCGCTCCAGTTG 2446061
QY 203 ThrGlnArgMetLeuCysAlaAsnGlnIleMetLeuLeuGlnThrPheProValPheGly 222
Db 2446062 ACCCAATCCAACTGTTACAGAGATTAGGGCGCCAGGGCATATTCACCGACGCTGCTG 2446121
QY 223 GlnGluProProValLeuValAspTrpLeuProTrpAsnHisIlePheGlyGlySerHis 242
Db 2446122 CGCAAGGTGAGCGCGCTGCTGTTTCTCTGCGCTACGATGTCGT-----GCCGCG 2446175
QY 243 AsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrLeuAspAspGlyLysProThr 262
Db 2446176 GCGATCGATGAGCGCGCTTCCAC-----TCCAAAGTCAC 2446211
QY 263 AlaGlnGlyPheAlaGluThrLeuArgAsn-----LeuSerGlnIleSerPro 278
Db 2446212 GTG---GGATTACACGACGACATCAAGATCTGCTGCCGATGTTGGGCGTTCAACCG 2446268
QY 279 ThrAlaTyrLeuThrValProLysGlyTrpGlnGluLeuValGlyAlaLeuGluArgAsp 298
Db 2446269 ACCGTGTGTGTGTCGCGCGGAGGTGTGTGAGAGGTTGTACAAACCGCGGACAGAAC 2446328
QY 299 SerThrLeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaA----- 316
Db 2446329 GCCGCCAACGCC-----GGCAAGGGCGAATCTTCGCGATCGCGCGCAGACC 2446376
QY 317 -----AlaGlyLeu-----Ser 320
Db 2446377 GCGGTGACTGAGCGAAGCTTGCGACCGCGCGGACCGGGGCTGTACTCGCGGCAAG 2446436
QY 321 GlnGlyIleTrpAspArgLeu----- 327
Db 2446437 CACGCGGTGTTCAACGCGTGTGCTACCGCAAGCTGCGGACCTGCGGCACTGC 2446496
QY 328 -----AspArgValAlaGluGlnHisCysGly 336
Db 2446497 CGCGCGCGCTCTCCGCGCGCGCGCGCTGCGGCGGCTGTGCTACTTCTATCCGCGC 2446556
QY 337 GluArgIleArgMetMetAlaGlyLeuGlyMetThrGlnThrAlaProSerCysThrPhe 356
Db 2446557 GCGCGTCTCACCATCTACGAGGATATGCTGAGCGGACCAATGGGGGCGCTCGCCATC 2446616
QY 357 ThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGluVal 376
Db 2446617 AGCCAGTTCAATGATCTAAAGATCGAGAACTGCGAAGCGCGGCCGCAACAGCTTA 2446676
QY 377 LysLeuValProValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSer 396
Db 2446677 CGCATGCCCGAC---GATGCGGAGCTGCTGTGCGC-----GGTGGCGTGTATTGAGC 2446727

QY 397 GlyTyrTrpArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCys 416
Db 2446728 GGCTACTGGCGCAACGAGCGGTATCACCGACGAGCATTC---ACCGACGCGCTGTTCAAG 2446784
QY 417 SerGlyAspAlaIleLysLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAsp 436
Db 2446785 ACCGTTGAT-----CTCGGTGCGGTGACCAAGACGGGTTCTTGACGATCAC 2446832
QY 437 GlyArgIleAlaGlnAspPheLysLeuSerSerGlyValPheValSerValGlyProLeu 456
Db 2446833 GGCGCGAAGAAAGAAATTATCGTACACCGCGGGGT-----AAAAATGCGCCCGCT 2446886
QY 457 ArgThrArgAlaValLeuGluGlyGlySerTyrValLeuAspValValAlaAlaPro 476
Db 2446887 GTGCTGAGAACCAAGCTCGCGGCCCACTCATGACGCCAGCGGTGTGTGGGAGC 2446946
QY 477 AspArgGluCysLeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSer 496
Db 2446947 GCCAAGCCTTCATCGCGCGTTAATC-----ACCATCGACCTGAGGCAATTGAG 2446997
QY 497 Gly-----LeuGlyLysGluAlaSerAsp 504
Db 2446998 GGCTGAGAGCAAGCAACAGCAAGACAGTGGCGCGTGGGTGGGCGATTGGCCACCGAC 2447057
QY 505 AlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuLysArgLeu 524
Db 2447058 CCGCATCTGATGCTCCGATCGACGCGGCC-----GTCAAAACAGGCC 2447099
QY 525 AsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThrPro 544
Db 2447100 AATCTT---GCGGTGTCAATGCTCGAGTCCGCAAGTTCCGAAATATCGCGTGCAC 2447156
QY 545 ProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnIleArgAlaValLeu 564
Db 2447157 TTACACGAGACACCGCGGAGCTGACCCGACATGAAGTCAACCAAGGTGTGTGCC 2447216
QY 565 GlnTrpArgSerAlaLysValAspAlaLeuTyrArgGlyGlu 578
Db 2447217 GAGAAAGTTGCTTCGATATCGAGGCGATCTCAACAGAGAA 2447258

RESULT 14
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FRIESCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 3,68e-13 Length: 4411529
Score: 259.00 Matches: 142
Percent Similarity: 37.92% Conservative: 106
Best Local Similarity: 21.71% Mismatches: 282
Query Match: 8.40% Indels: 124
DB: Gaps: 26

US-09-750-986d-30 (1-589) x US-09-103-840A-1 (1-4411529)

```

Db      425 TCRAAATTAATACTATGAACTATATATATAGACTTAATGAAGACTAGAGCT 484
Qy      170 TLeGluAlaAspAenAlaPheAlaAlaThrGlyProAspThr----- 183
Db      485 TATCAATGCTCAACAACCTTATTTCTCAAAATTCCTAGTAGTATCGATGAAGAAA 544
Qy      184 -----TLeAlaAspPheLeuPheThrSergly 192
Db      545 TTTAAACATATCTTTTAATCGAGACGATCGAGTTGATTAATGTTTCTTCGTGT 604
Qy      193 SerThrLeuLeuProGlyAlaValProThrThrGluMetLeuCyAlaAenGluAla 212
Db      605 ACAACTGCTGCGGAAGGAGTCATGCTAATCACAAGAAATATGTGCA-----CGA 658
Qy      213 MetLeuLeuGlnThrPheProValPheGly-----GluGluProProValLeuVal 229
Db      659 TTTTCTCTTGCAAAAGATCTTACTTTGTGTAACGCAATTATCCACGACGCAATTTTA 718
Qy      230 AspThrLeuProThrAsnHisThrPheGlyGlySerHisAsnIleGly----- 245
Db      719 ACGGTATACCTTCCACCATGGTTTGTGATGATGACACCATTAAGATCTTACTTGT 778
Qy      246 -----TLeValLeuTyraAsnGlyGlyThrTyrrTyraAspAspGlySProThr 262
Db      779 CGATTCGAGTTGTTCTTAATGCAC----- 802
Qy      263 AlaGlyGlyPheAlaGluThr-----LeuArgAsnLeuSerglyIleSerpThrAla 280
Db      803 -----ACGTTGAGAAAACATATTTCTCAATCATTAACAGATTAATAAGTCGAAGT 856
Qy      281 TyrrLeuThrValProGlySgIlyTrpGluGluLeuValGlyAlaGluGluAspSerpThr 300
Db      857 ACTTATCTGTACCA-----ACATTAATGCGCATTTCTTGCAAAAAGTCATTA 904
Qy      301 LeuArgGluArgPhePheAlaArgMetLeuLeuPhePheAlaAlaAlaGlyLeuSer 320
Db      905 GTTGAAAGATGACATTATGACACTTAAGAAATTCATCGTGTGCGCACCTTATCA 964
Qy      321 GlnGlyIleThrAspArgLeuAspArgValAlaGluGlnHisCySgIlyGluAlaArg 340
Db      965 AAGAAAT-----GGGAGATGCTGAAA 988
Qy      341 -----MetMetAlaGlyLeuGlyMetThrGluThrAlaProSer 353
Db      989 AAACGGTTTAATTAACCTTGTACGCAAGGATGATTAACAGAAACCACTCGGCT 1048
Qy      354 CySerThrPheThrThrGlyProLeuSermetalGlyTyrrIleGlyLeuProAlaProGly 373
Db      1049 GTTTTATTAACACCGAAAGGTGACGCCAAACCGGATCACTGTTAAATAGTACATTT 1108
Qy      374 CySgIlyValIleValLeuVal-----ProValAspGlyIlyLeuGluGly----- 387
Db      1109 CAGCGTTTAAAGTTGTCGATCTTACCAACAGAAAATTTTGGGGCCMAATGAACCTGA 1168
Qy      388 -----ArgPheHisGlyProHisValMetSerglyTyrrTrpArgAlaProGluGlnAsn 405
Db      1169 GAATTTGATTTTAAAGCCCGATGATATGAAAGGTTATTAATATGAAGAACTACT 1228
Qy      406 AlaGlnAlaPheAspGluGlyGlyTyrrCySerglyAspAlaIleIleuSleuAlaAsp 425
Db      1229 AAGGCAATTATGATATGACGAGTGTGCGCTGTGAT--ATTGCTTATTAAGAC 1285
Qy      426 ProAlaAspProGlnIlySgIlyLeuMetPheAspGlyValGlyIleAlaGluAspPheIlySleu 445
Db      1286 AAT-----GATGGCCATTTTATATATGATGACAGGCTG 1318
Qy      446 SerSerglyVal-----PheValSerValGlyProLeuArgThrArgAlaValIleu 462
Db      1319 AAGTCACTGATTAATTAAGGTTATCAAGTTGACCTGCTGAATTTGAGGAATATCTC 1378
Qy      463 GluGlyGlySerTyrrValLeuAspValValValAlaAlaPheAspArgGlyCySleuGly 482

```

```

Db      1379 TTACAATCCGATATATTGTTGATCCGGCGCTTACTGTATACCGGATGAAGCCGGCG 1438
Qy      483 LeuLeuValPheProArgLeuLeuAspCyArgArgAlaLeuSerglyLeuGlyIlySgIlyAla 502
Db      1439 -----GAGCTTCC-ACGTCAGGTGTGTAGT----- 1464
Qy      503 SerAspAlaGlyValLeuAlaSerGluProValArgAlaThrPheAlaAspTrp----- 520
Db      1465 -----ACAGACTGGAAATA 1479
Qy      521 LeuTyraArgLeuAsnArg-----GluAlaThrGlyAsnAlaSerArgIle 535
Db      1480 TCRAACGAACAATGCTACAGATTAATGTGCGCAATCAAGTTCAACAGCCAAATGGCT 1539
Qy      536 MetTrpValGly 539
Db      1540 ACGTGTGGGCT 1551

RESULT 13
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 3,67e-13 Length: 4403765
Score: 259.00 Matches: 142
Percent Similarity: 37.92% Conservative: 106
Best Local Similarity: 21.71% Mismatches: 282
Query Match: 8.40% Indels: 124
DB: 3 Gaps: 26

US-09-750-986D-30 (1-589) x US-09-103-840A-2 (1-4403765)
Qy      2 ArgSerLeuGluAlaLeuLeu-----ProPheProGlyArgIleLeu 15
Db      2445438 AGCTCAAGTAGAGGTGTGATCGTGGCGTGAATTAAGGTCGCCGCCCATTCACATGCGGC 2445497
Qy      16 GluArg-----LeuGlnHisTrpAlaIlyThrArgProGluGln 28
Db      2445498 GAGCAGACGACCACTCCGCCGATGCTGTTGAGACAT--GAACGTACAGATCCCGACATAC 2445554
Qy      29 ThrCyValAlaAlaArgAlaAlaAenGlyGluTrpArgArgIleSerpTrpAlaGluMet 40
Db      2445555 GTC--ATCATCAAGCTGATGACGACGCGGTGACCGAATGTCACGTTGGGAGGCA 2445611
Qy      49 PheHisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrrIlyLeuSerAlaGluArg 68
Db      2445612 GCCAACGAGATTCGTCGCCGCGGCTTCGCGTTGATTTACCTGGCGGCTGACGCGCGCAT 2445671
Qy      69 ProLeuLeuIleValSerglyAsnAspLeuGlnHisLeuGlnLeuAlaPheGlyAlaMet 88
Db      2445672 CGGATGATCATCTTCTTCTGCAACCCGCTACGAGTGGCGATCTCGATTTCCGGAATTCG 2445731

```

Db 545 TTAAACCTATTCTTTAATCGAGACGATCGTGGTGAATGATTTCTTCGTGT 604
Qy 193 SerThyLeuProLysAlaValProThrThrglnMetLeuCysAlaAsnGln 212
Db 605 ACAACTGCTGCGGAAAGAGTCACTTAACCAAGAAATATTGTGCA-----CGA 658
Qy 213 MetLeuGlnThrPheProValPheGly-----GluGluProProValLeuVal 229
Db 659 TTTCTATTGCAAAAGATCTTACTTTGTGTACCAATTAATCCACGTGACAAATTTTA 718
Qy 230 AspThrLeuProTrpAsnHisThrPheGlyGlySerHisAsnIleGly----- 245
Db 719 ACGGTAAATACCTTCCACCAATGTTGTGTATGATGACCAATTAAGTACTTACTGT 778
Qy 246 -----IleValLeuTyraAsnGlyGlyThrTyrrTyrrLeuAspArgLysLeuProThr 262
Db 779 GGATCCGAGTGTGTTCTAATGCAC----- 802
Qy 263 AlaGlnGlyPheAlaGlnThr-----LeuArgAsnLeuSerGluIleSerProThrAla 280
Db 803 -----ACGTTGAAGAAACATATTCTCAATCAATTAACAAGTTAATAAGTGGAAGT 856
Qy 281 TyrLeuThrValProLysGlyTrpGluGlnLeuValGlyAlaLeuGluArgAspSerThr 300
Db 857 ACTTACTTGACCA-----ACATTAATGCGCATTTCTTGCAAAAGTGCATTA 904
Qy 301 LeuArgGlnArgPhePheAlaArgMetLysLeuPhePheAlaAlaAlaGlyLeuSer 320
Db 905 GTTGAAGAAAGTCACTTAATGCGACTTAAGAAATTCGATCGTGGCGACTTATCA 964
Qy 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGlnGlnHisCysGlyGluArgIleArg 340
Db 965 AAGAAATTT-----GGGAGATGCTGAAA 988
Qy 341 -----MetMetAlaGlyLeuGlyMetThrGluThrAlaProSer 353
Db 989 AAACGGTTTAATTAACCTTTGTCAGGCAAGGATGATGAATTAACAGAAACACTTCGCT 1048
Qy 354 CysThrPheThrThrglyProLeuSerMetAlaGlyTyrrIleGlyLeuProAlaProGly 373
Db 1049 GTTTTAATTAACCGAAAGGTGACGCAAAACGGGATCAACTGTTAAATTAAGTACATTA 1108
Qy 374 CysGluValLysLeuVal-----ProValAspGlyLysLeuGlnGly----- 387
Db 1109 CAGCGCTTAAAGTTGTCGATCCTACAAACGAAAAATTTGGGCGCAATGAACCTGCA 1168
Qy 388 -----ArgPheHisGlyProHisValMetSerGlyTyrrTrpArgAlaProGluGlnAsn 405
Db 1169 GAATTCATTTTAAAGCCGATGATATGAGGGTATTATATATATGAAGAACTACT 1228
Qy 406 AlaGlnAlaPheAspGluGlnGlyTyrrTyrrCysSerGlyAspAlaIleLysLeuAlaAsp 425
Db 1229 AAAGCAATTAATTAATTAAGTATGCGATGGTGGCTGTATAT--ATTGCTTATTATGAC 1285
Qy 426 ProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLysLeu 445
Db 1286 AAT-----GATGCGCATTTTATATATGATGCGACAGGCTG 1318
Qy 446 SerSerGlyVal-----PheValSerValGlyProLeuArgThrArgAlaValLeu 462
Db 1319 AAGTCACTGATTAAATTAAGTTATCAAGTTGACCTGCTGAATTAAGGGAATATCTC 1378
Qy 463 GluGlyGlySerTyrrValLeuAspValValValAlaAlaProAspArgGluCysLeuGly 482
Db 1379 TTACAAATCCGATTAATGATGATGCGGCGTTACTGTTATCCGATGAAGCGCGGCG 1438
Qy 483 LeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGluAla 502
Db 1439 -----GAGCTTCC-AGCTGCAAGGCTGTGAT----- 1464
Qy 503 SerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp----- 520

Db 1465 -----ACAGACTGGAATA 1479
Qy 521 LeuLysArgLeuAsnArg-----GluAlaThrGlyAsnAlaSerArgIle 535
Db 1480 TCTAAACGAACAAATCGTATCAAGATTAATGTTGCCAGTCAAGATTCAACAGCAATGCGT 1539
Qy 536 MetTrpValGly 539
Db 1540 ACGTGTGGGGT 1551

RESULT 12
US-09-396-154-11
Sequence 11, Application US/09396154
Patent No. 6602677
GENERAL INFORMATION:
APPLICANT: Wood, Keith V.
APPLICANT: Hall, Mary P.
TITLE OF INVENTION: Thermostable luciferases and methods of
FILE REFERENCE: 341.012US1
CURRENT APPLICATION NUMBER: US/09/396,154
EARLIER FILING DATE: 1999-09-15
EARLIER APPLICATION NUMBER: US 09/156,946
EARLIER FILING DATE: 1998-09-18
EARLIER APPLICATION NUMBER: PCT/US98/19494
EARLIER FILING DATE: 1998-09-18
EARLIER APPLICATION NUMBER: US 60/059,379
NUMBER OF SEQ ID NOS: 93
SOFTWARE: Fasteq for Windows Version 3.0
SEQ ID NO 11
LENGTH: 1639
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Mutant luciferase
US-09-396-154-11

Alignment Scores:
Pred. No.: 3,37e-19 Length: 1639
Score: 262.50 Matches: 129
Percent Similarity: 36.52% Conservative: 77
Best Local Similarity: 22.87% Mismatches: 218
Query Match: 8.52% Indels: 141
DB: 4 Gaps: 23

US-09-750-986D-30 (1-589) x US-09-396-154-11 (1-1639)
Qy 30 CysValAlaAlaAlaArgAlaAlaAsnGlyGluTrpArgArgIleSerTyrrAlaGluMetPhe 49
Db 119 TGCATGCAATGTCGCAAAATGCTCATACAAA-----GAAATGTTTATATGAAAGTTTCG 175
Qy 50 HisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrrGlyLeuSerAlaGluArgPro 69
Db 176 AAACGTGCTGCTGTTTACGCGGAAAGTTTAAAGATATGATTAACAAACAAACGACACA 235
Qy 70 LeuLeuIleValSerGlyAsnAspLeuGlnHisLeuGlnLeuAlaPheGlyAlaMetTyrr 89
Db 236 ATAGCGGTGTGAGCAAAATGGTGTGCAATTTTCTTCCTGTAATTCGATCATTTGTAT 295
Qy 90 AlaGlyIleProTyrrCysProValSerProAlaTyrrSerLeuLeuSerGlnAspLeuAla 109
Db 296 CTTGGAATTAATTTGGCACTGTTAAACGATTAATAC-----ATTGAACGTGAATTAATA 349
Qy 110 LysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAspAlaAla 129
Db 350 -----CACAGCTCTGCTATGTAAACCAACGATGTTTGTTCCTCAAGAAATACT 400
Qy 130 ProPheGlnArgAlaIleGluThrIleLeuProAspAspValProAlaIlePheThrArg 149
Db 401 -----TTTCAAAAGTACGTAAT-----GTAATA 424
Qy 150 GlyGluLeuAlaGlyLysArgThrValSerPheAspSerLeuLeuGlnProGlyGly 169

```

Db      659 TTTTCTTGCAAAAGATCTTACTTTGGTAACGCAATTATCCACGACGCAATTTTA 718
Qy      230 AepTrrpLeupProTrrpAenhiStrhpheGlySerhiAenileGly----- 245
Db      719 ACGGTAAATACCTTCCACCATGTTTGGTATGATGACCACTAGGACTTACTTACTGT 778
Qy      246 -----lLeValLeuTyrrAenglyGlyThrTyrrLeuAepAepGlyLysProThr 262
Db      779 GGATTCGAGTGTGTCTTAATGAC----- 802
Qy      263 AlaGlnGlyPheAlaGlnThr-----LeuArgAenLeuSerGluLLeSerProThrAla 280
Db      803 -----ACGTTTGAAGAAAACTATTTCTTACATCATTAACAAGTTTAAGGGAAGT 856
Qy      281 TyrrLeuThrValProlySgLyTrpGluGluLeuValGlyAlaLeuGluArgAepSerThr 300
Db      857 ACTTTACTGTGACCA-----ACATTTAATGCGATTTCTTGCAAAAAGTGCACTTA 904
Qy      301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSer 320
Db      905 GTTGAAAAGTACATTTATGACACTTAAAGAAATTGCACTGTGCGCGCACTTATATCA 964
Qy      321 GlnGlyLleTrpAepArgLeuAepArgValAlaGluGlnhiScySgLyGluArgLLeArg 340
Db      965 AAAGAAATT-----GGCGAGATGCTGAAA 988
Qy      341 -----MetMetAlaGlyLeuGlyMetThrGluThrAlaProSer 353
Db      989 AAACGGTTAAATTAACCTTTGTCAGGCAAGGATGATTAACAGAAACCACTTCGGCT 1048
Qy      354 CyethrPheThrThrGlyProLeuSerMetAlaGlyTyrrLleGlyLeuProAlaProGly 373
Db      1049 GTTTTAATTAACCCGAAANNNNNNCCAGACCGGAGTCAACCTGTAATAATAGTACATTT 1108
Qy      374 CySgLyValLysLeuVal---ProValAepGlyLysLeuGlnGly----- 387
Db      1109 CAGCGTGTAAAGTGTGCGATCTTACAAACGAAAAATTTTGGGCCCAATGAACCTGGA 1168
Qy      388 -----ArgPhehiSgLyProhiSvalMetSerGlyTyrrTrpArgAlaProGluGlnA 405
Db      1169 GAATGTATTTTAAAGCCCGCATGATATGAGGGTATTAATTAATGAAGAACTACT 1228
Qy      406 AlaGlnAlaPheAepGluGlnGlyTyrrCySerGlyAepAlaLleLysLeuAlaAep 425
Db      1229 AAAGCAATTATTAATGAAGCGATGCGCTGCTGTGATATTT----- 1273
Qy      426 ProAlaAepProGlnLysGlyLeuMetPheAepGlyArgLleAlaGluAepPheLysLeu 445
Db      1274 -----GCTTATTAATGAC----- 1285
Qy      446 SerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlnGlyGly 465
Db      1286 AATGATGCGCATTTTATTTATTTGTCAGACGCTGAAGTCATTAATTAATTAAGT--- 1342
Qy      466 SerTyrrValLeuAepValValAlaAlaProAepArgGluCySLeuGlyLeuLeuVal 485
Db      1343 -----TATCAGGTTCACCTGCTGAATTAAGGGAATA---CTCTTAACA 1384
Qy      486 PheProArgLeuLeuAepCyArgAlaLeuSerGlyLeuGlyLysGluAlaSerAaspAla 505
Db      1385 CATCCGATATTTGTGTATGTCG---GGCGTACTGCTATTCGCAATGAAGCCGGCGGAG 1441
Qy      506 GluValLeuAlaSerGluProValArgAlaLleTrpPheAlaAepTrpLeuLysArgLeuA 525
Db      1442 CTTCAGACTCAGGTGTGTGATGACACACTGGA-----AATATCTTAAC 1486
Qy      526 ArgGluAlaThrGlyAep-----AlaSerArgLleMetTrp----- 537
Db      1487 GAACAAATCGTACAAATTTTGTTCAGTCACATTCACAGCAAAATGCTACGTGCT 1546
Qy      538 ---ValGlyLeuLeuAepThrProProSerLLeAepLysGlyValAlaThrAepLys 555

```

```

Db      1547 GGGGTAAATTTTGGATGAATTCACCAAGATCACTGCAAAAATTAATGACAGAAAA 1603
RESULT 11
US-09-396-154-43
; Sequence 43, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Mood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable Luciferases and methods of
; TITLE OF INVENTION: Production
; FILE REFERENCE: 341.012051
; CURRENT APPLICATION NUMBER: US/09/396,154
; EARLIER FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: US 09/156,946
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FaSeq for Windows Version 3.0
; SEQ ID NO 43
; LENGTH: 1639
; TYPE: DNA
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Mutant Luciferase
US-09-396-154-43

Alignment Scores:
Pred. No.:      2,61e-19      Length:      1639
Score:          263.50      Matches:      128
Percent Similarity: 36.52%      Conservative: 78
Best Local Similarity: 22.70%      Mismatches: 218
Query Match:      8.55%      Indels:      141
DB:              4          Gaps:      23

US-09-750-986d-30 (1-569) x US-09-396-154-43 (1-1639)
Qy      30 CySvalAlaAlaArgAlaAlaAenglyGluTrpArgArgLLeSerTyrrAlaGluMetPhe 49
Db      119 TGCATGCAATGTCACAAATGCTCTCATACAAAA---GAAAAAGTTTATATATGAGGTTTCTG 175
Qy      50 HiAasnValArgAlaLleAlaGlnSerLeuLeuProTyrrGlyLeuSerAlaGluArgPro 69
Db      176 AAACGTGCTGTCGTTTACCGGAAAGTTTAAAGTAAGATGATTAACAAAAACGACACA 235
Qy      70 LeuLeuLleValSerGlyAenAepLysGluhiSLeuGlnLeuAlaPheGlyAlaMetTyrr 89
Db      236 ATAGCGGTGTGTGCGCAAAATAGCTTCGAATTTTCTCCTGTATTTGCAATGATGTAT 295
Qy      90 AlaGlyLleProTyrrCySProValSerProAlaTyrrSerLeuLeuSerGlnAepLeuAla 109
Db      296 CTTGGAATATATTTGGCACCTGTTTAACGATTAATATC-----ATTGAACGTGAATTAATA 349
Qy      110 LysLeuArgHisLleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAepAlaAla 129
Db      350 -----CACGTGTGTGATTTGTAACCCGCAATGTTTGTTCCTCAAGAAATACT 400
Qy      130 PheGlnAlaArgAlaLleGluThrLleLeuProAepAepValProAlaLlePheThrArg 149
Db      401 -----TTTCAAAAATGACTGAAT-----GTAAAA 424
Qy      150 GlyGluLeuAlaGlyArgThrValSerPheAepSerLeuGlnGluGlnProGlyGly 169
Db      425 TCTAAATTAATAATCTATTTGAACTATTAATTAATTAATTAATGAAGACTTGAAGAGGT 484
Qy      170 LLeGluAlaAepAenAlaPheAlaLleThrGlyProAepThr----- 183
Db      485 TATCAATGCTCAACAACCTTATTTCTCAAAATTCGATAGTAAATGTCGACGTAATAAAAA 544
Qy      184 -----lLeAlaLysPheLeuPheThrSerGly 192

```


QY 319 LeuSerGlnGlyIle-----TTP-----AspArgIleuAspArg 329
Db 991 CGTTCAATTAGCGCGTATCTCAATGCTTATCGCTTTCGCGATAAAGATCTTCAA 1050
QY 330 ValAlaGluGlnHisCysGlyValArgIleArgMetMetAla----- 343
Db 1051 CCACATACGTCAAAATTTAGAGTGGCGAGTCTTTTACCTCCGCGAGCGCTCGTTTA 1110
QY 344 -----GlyLeuGly 346
Db 1111 GATGATGCGGTGATCGGCTTTTCTTACGCTGCTATTAATAATTAATATGTTAGGC 1170
QY 347 MetThrGluThrAlaProSerCysThrPheThrThryGlyProLeuSerMetAlaGlyTyr 366
Db 1171 ATGACGAGAGACTGTGTCCACGCGATCTTGTGGAGAAACCATTTAATAAGGCTCT 1230
QY 367 IleGlyLeuProAlaProGlyCysGlyValLysLeuValProValAspGlyLysLeuGlu 366
Db 1231 ATTGGTACTCCGTTACCGGGGGTTGAGTACGT-----ATAGCGGAAGAAATGAG 1281
QY 387 GlyArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAla 406
Db 1282 ATCCAGTGGCGGTAGTATCGTATCGATGAAGGTTATTTAATAAGCTGAGAGACAGT 1341
QY 407 GlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleLysLeuAlaAspPro 426
Db 1342 GCTGCGTTTACTGAAGATGCTGTGGTACGTACCGGTATCGGGTGCGATTA----- 1392
QY 427 AlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLysSer 446
Db 1393 --GATGAGATGGGATGTTATATTATCCGAACGTTTGAAGAAGTCTTAAGAAAACCTCA 1449
QY 447 SerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGluGlyLysSer 466
Db 1450 AATGGAATAATATGTA-----CCGCAATGATTTGAAGTACATTAGGCGAAGATCGC 1503
QY 467 TyrValLeuAspValValValAlaAlaProAspArgGluCysLeuGlyLeuValPhe 486
Db 1504 TTATATGAGCATATCGCTGTGATTTGACATATCCGTAATTTGTTCTGCACTGATCGTA 1563
QY 487 ProArgLeuLeuAspCys-----ArgAlaLeuSerGlyLeu 498
Db 1564 CCT-----TGCCTTGATGCTTAAGAAACATCGTCGAGCATTTAAAT----- 1605
QY 499 GlyLysGlnLysSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAla 518
Db 1606 TTTAAATATACATCGCATAGAGATTATTCGCCATATCTAAGATCAAGAATTTTGGAT 1665
QY 519 AspTrpLeuLysArgIleuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpVal 538
Db 1666 GAACGCTGTGGTGAATATGCAACGTAAAT---TTGCTAGTTCATCATCAAGTTAAACGATTT 1722
QY 539 GlyLeuLeuAspThrProProSerIleAspLysGlyGluValThrAspLysGlySerIle 558
Db 1723 ACCCTATTCGGAAGGGTTTACGATGAGTCTGGAATATGACA----- 1767
QY 559 AsnGlnArgAlaValLeuGlnTrpArgSerAlaLysLeuAspAlaLeuTyrArgGlyGlu 578
Db 1768 -----CTTACCTTTAAATTAAGACGTAAATAATTTCTGACGCTATCGAATGAA 1818
QY 579 AspGlnSerMetLeuArgAsp 585
Db 1819 ATGAGCAGATGTATCAAGAG 1839

RESULT 10
US-09-396-154-10
; Sequence 10, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; TITLE OF INVENTION: production

FILE REFERENCE: 341.012US1
CURRENT APPLICATION NUMBER: US/09/396,154
CURRENT FILING DATE: 1999-09-15
EARLIER APPLICATION NUMBER: US 09/156,946
EARLIER FILING DATE: 1998-09-18
EARLIER APPLICATION NUMBER: PCT/US98/19494
EARLIER FILING DATE: 1998-09-18
EARLIER APPLICATION NUMBER: US 60/059,379
EARLIER FILING DATE: 1997-09-19
NUMBER OF SEQ ID NOS: 93
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 10
LENGTH: 1639
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: unsure
LOCATION: (1067)..(1072)
OTHER INFORMATION: Unknown nucleotides
OTHER INFORMATION: Mutant luciferase
US-09-396-154-10
Alignment Scores:
Pred. No.: 2,61e-19 Length: 1639
Score: 263.50 Matches: 132
Percent Similarity: 37.31% Conservative: 84
Best Local Similarity: 22.80% Mismatches: 226
Query Match: 8.55% Indels: 137
DB: Gaps: 24
US-09-750-986D-30 (1-589) x US-09-396-154-10 (1-1639)
QY 30 CysValAlaAlaArgAlaAlaAsnGlyGluTrpArgArgIleSerTyrAlaGluMetPhe 49
Db 119 TGCATGCAATGTCGAATGCTCATACAAA---GCCCGTGTATATAGAGGTTTTA 175
QY 50 HisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeuSerAlaGluArgPro 69
Db 176 AATGTGTCGTGCTTAAAGGGAAGTTTAAAGATATGATTAAGATTAACAAACGACACA 235
QY 70 LeuLeuIleValSerGlyPheAspLeuGlnHisLeuGlnLeuAlaPheGlyAlaMetTyr 89
Db 236 ATGCGGTGTGTCGCAAAATGGTTGCAATTTTCTCTCTGTAATGTCATCATTTGAT 295
QY 90 AlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeuSerGlnAspLeuAla 109
Db 296 CTGGAAATATTTGGCACCCTGTTAACGATAATAC-----ATGACCTGAATTAATA 349
QY 110 LysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAspAlaAla 129
Db 350 -----CACAGTCTGGTATTTGTAACCCAGCATAGTTTTTGTCTCCAGAAATACT 400
QY 130 ProPheGlnArgAlaIleGluThrIleLeuProAspAspValProAlaIlePheThrArg 149
Db 401 ---TTTCAAAAGTACTGAAT-----GTAATA 424
QY 150 GlyLysLeuAlaGlyLysArgThrValSerPheAspSerLeuLeuGlnProGlyGly 169
Db 425 TCTTAATTTAAATCTGTAAGAACTATTTATTTAGACTTAATGAAAGACTTACGAGGT 484
QY 170 IleGluAlaAspAsnAlaPheAlaAlaThrGly----- 180
Db 485 TATCAATGCTCTCAACAACCTTATTTCTCAAAATTCGATATTAATCTGACGTAATAAAA 544
QY 181 -----ProAspThrIleAlaLysPheLeuPheThrSerGly 192
Db 545 TTTAAACATATTTCTTTATTCGACAGCATACAGGTGCTGTAATTTGTTCTTGTGT 604
QY 193 SerThrLysLeuProLysAlaValProThrThrGlnArgMetLeuCysAlaAsnGlnGln 212
Db 605 ACAACTGCTGTCGGAAGGAGTCAATGCTTAATCTACACAAAGATTTGTGTGA-----CGA 658
QY 213 MetLeuLeuGlnThrPheProValPheGly-----GluGluProProValLeuVal 229

```

QY 355 -----ThrpheThrThrglyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaPro 372
DB 1774 TGGCATGATTTCCAAATTAAACCAAAATTC-----ATCGCACACGATGCCA 1821
QY 373 GLYCYGGLValIysLeuValProValAspGlyIlyLeuGluGlyIlePheHISGlyPro 392
DB 1822 AAAGCGAAATGAAA-----ATTGGGGAATAATGAAATCCCTTGTGCGTGGCGGA 1872
QY 393 HlValMetSerGlyTyrTrpArgAlaProGluGlnAsnIleGlnAlaPheAspGluGlu 412
DB 1873 ATGCTGATGAAGGCTTTTACAGAACCCAGAGAAACGCGTCAAGCTTCAACGAAGAT 1932
QY 413 GLYTyrTrpCysSerGlyAspAlaIleIysLeuAlaAspProAlaAspProGlnIlyGly 432
DB 1933 GGTTCCTTAAAACTGGCGCATGCGAGAAATTT-----GACGAACAAGCGCAAT 1980
QY 433 LeuMetPheAspGlyArgIleAlaGluAspPheIysLeuSerGlyValPheValSer 452
DB 1981 TTAATTATTAATCCATCGTATCAAGAAATTAATGAAACCTCAACGCGAAATATATCGCA 2040
QY 453 ValGlyProLeuArgThrArgAlaValIleuGluGlyIlySerTyrValLeuAspValVal 472
DB 2041 -----CCACAAATATATCGAAAGCAAAATCGTAAGATTAATTTATCGAAACAAATTCG 2094
QY 473 ValAlaAlaProAspArgGlyCysLeuGlyLeuLeuValPheProArgLeuLeuAspCys 492
DB 2095 ATCATGCGTATGCGRAAAATATGATTCGCGCTTATTGTGCT-----TGC 2142
QY 493 ---ArgAlaLeuSerGlyLeuGlyIlyGluAlaSer-----AspAlaGlu 506
DB 2143 TTGTATAGTTTGAAGAAATACGCTAAACAGCTCAATTTAAATATCATGACCGTTAGAA 2202
QY 507 ValIleuAlaSerGluProValArgAlaIlePheAlaAspTrpLeuIysAspGluAsnArg 526
DB 2203 CTACTAAAAAATCTGACATTCGAAATGTTGAGCATGCTATTAATGCGGTGCAAAA 2262
QY 527 GluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspTrpProposer 546
DB 2263 GAA---TTGGCTCATTTTCGAGCAAGTAAATAATTCACGTTACTTCTCAAGCATTCAGC 2319
QY 547 IleAspIysGlyGluValThrAspIysGlySerIleAsnGlnArgAlaValIleuGlnTrp 566
DB 2320 ATTAAATTAGCGCAAAATTACACCAACATTAAATTAAGTAGAAAGATTTTGAACGT 2379
QY 567 ArgSerAlaIysValAspAlaLeuTyrArgGlyGluAsp 579
DB 2380 TATCGACGAATATGAAATGATATCATTTCAAGAA 2418

RESULT 9
US-09-543-681A-2894
; Sequence 2894, Application US/09543681A
; Patent No. 6605708
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2894
; LENGTH: 1845
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2894

Alignment Scores:
Pred. No.: 1,66e-19 Length: 1845
Score: 266.00 Matches: 140
Percent Similarity: 37.87% Conservative: 105
Best Local Similarity: 21.64% Mismatches: 270

```

```

Query Match: 8.63% Indels: 132
DB: 4 Gaps: 23
US-09-750-986D-30 (1-569) x US-09-543-681A-2894 (1-1845)

QY 14 IleLeuGlnArgLeuGlnIleStrpAlaIysThrArgProGluGlnIleThrCysValAlaAla 33
DB 70 ATATATTAGAAAGATTGAGATGCAATTTTCATCTTCTGGGGAAAAAAGCGCTTATCGTCAG 129
QY 34 ArgAlaAlaAsnGlyGluTrpArgGlyIleSerTyrAlaGluMetPheHISAsnValArg 53
DB 130 TGGAGGCTTAAGCTTAA---ATCGCATGTCTAGGCGCTGACTACCAAGAAACACGT 186
QY 54 AlaIleAlaGlnSerLeuLeuProTyrGlyLeuSerAlaGluArgProLeuIleVal 73
DB 187 GCACCTCAATGATTAATCTCATATGCGCGTGGCGCTTCAAGAAATGCGGATTTT 246
QY 74 SerGlyAsn-----AspLeuGlnIleLeuGlnIleuAlaPheGly 86
DB 247 TCCCAAAAGCATGTGATTGCTATGCTATGCTGATATGCTACTTGCAGTTGGCT----- 300
QY 87 AlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeuSerGln 106
DB 301 -----GCTGTCAAGTTCCTTTATTCACCAAGTACTGTG 336
QY 107 AspLeuAlaIysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAla 126
DB 337 GAACAAAGCC-----GCTTATATCATCAACGATCGGATTTGCTTTTGTGGT 390
QY 127 AspAlaAlaProPheGlnArgAlaIleGlu----- 136
DB 391 GACCAAAAGAAATATGACGTTGTTCTGAGTTAAGCGCTATTATGCCGCACTTAACAT 450
QY 137 -----ThrIleLeuProAspAspValProAlaIlePheTrpArg 149
DB 451 ATTATGTGTTAATTCCTCAGGTCGTTTAAATCCAAACATACCTTTATCATCATGACA 510
QY 150 GlyGluLeuAlaGlyArgArgThrValSerPheAspSerLeuLeuGlnIleProGlyGly 169
DB 511 AACTTATATCAGTACGTCAATCT---CAATATGACTCTCTGTCACAA----- 558
QY 170 IleGluAlaAspAsnAlaPheAlaAlaThrGlyProAspThrIleAlaIysPheLeuPhe 189
DB 559 ---CGTATGAGAAATGTTGTTA-----GATGATTTGTTTACAGCATTTAT 603
QY 190 ThrSerGlySerThrIysLeuProIyAlaValProThrThrGlnArgMetLeuCyAla 209
DB 604 ACTCGGTGTAACAACAGGTGAGCTTAAAGGTGTTATGCTGATTAACCAACATCGGCTCT 663
QY 210 AsnGlnIleMetLeuLeuGlnIleThrPheProValPheGlyGluGluProProValLeuVal 229
DB 664 -----CAACTTATATGATGATGATGACGCCCTTTCTATTATCGGATTAAGATGTTTCATTA 717
QY 230 AspTrpLeuProTrpAsnHisThrPheGlyIysSerHisAsnIleGlyIleValLeuTyr 249
DB 718 TCGTTTATACGTTATACATGATTTTGAACGGCATGAGCTTCAATGTTATGCTACT 777
QY 250 AsnGlyGlyThrTyrIleuAspArgGlyIysProThrAlaGlnIleGlnIlePheAlaGluThr 269
DB 778 GGGGCTATTATGTTATCTCACAGTACCCTATGCCGTAAGGAAGCGATGCT----- 831
QY 270 LeuArgAsnLeuSerGluIleSerProThrAlaTyrIleuThrValProIysGlyTrpGlu 289
DB 832 -----GCAGTAAACCAACCGTTATGCTGACGTGCAAGCTTTTATCGAA 876
QY 290 GluLeuValAlaIleuGlnArgAspSerThrLeuArgGluArgPhePheAlaArgMet 309
DB 877 AAAGTATATAGGCGCATTCACAGAAAGTTTCTCAAGACCTCGACTA-----CGCCAA 930
QY 310 IysLeuPhePhePheAla-----AlaAlaGly 318
DB 931 TCGATGTTTAAATGGGCTCTTAAACAGGGTGAAAAACAAACGTCAAGCACATTACTGTA 990

```

```
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (122336) ..(122336)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (131340) ..(131340)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (131360) ..(131360)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (139910) ..(139910)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (140398) ..(140398)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (142750) ..(142750)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145058) ..(145058)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145171) ..(145171)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (147197) ..(147197)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145942) ..(145942)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (150841) ..(150841)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (152500) ..(152500)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature

Alignment Scores:
Pred. No.: 1.76e-16 Length: 1830121
Score: 283.00 Matches: 137
Percent Similarity: 38.44% Conserved: 114
Best Local Similarity: 20.98% Mismatches: 252
Query Match: 9.18% Indels: 150
DB: 4 Gaps: 28

US-09-750-986d-30 (1-589) x US-10-329-960-1 (1-1830121)
QY 15 leuGluaRgLeuGluhiSTpAlaYvThrARgProGluGlnThrCySValaIaAarg 34
DB 646 GTTCATGTAATTCACAAACAAACCAAACTGTAACAATATGACCGCACTTCCTATAA 705
QY 35 AlaAlaAsnGlyGluTrpARgArgIleSerTyraIaGluwecPheHisAenValArgAla 54
DB 706 ---GAACACGGCTTGGCGCAAGATATCTCTTGAAAAAATTCAAGACCACTCAATCAA 762
QY 55 lIeAlaGlnSerLeuLeuProTyrglyLeuSerAlaGluArgProLeuLeuIleValSer 74
DB 763 CTTTCTCGAGCATTCCTGCTGCTACAAATATTAATGACGATTAATAATTCGCAATTTTGGC 822
```

```
QY 75 GlyAsn-----AspLeuGluhiSTeugLInleuAlaPheGlyAla 87
DB 823 CATATATATGAAACGTTGCAACATCGTTGACATTCGACCTTACAAATTCGA----- 873
QY 88 MetTyraIaGlyIleProTyrgyProValSerProAlaTySerLeuLeuSerGlnAsp 107
DB 874 -----GCATTCACAGTCCTATTATTAACCAACCAATACGACCCGAC 912
QY 108 LeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaIaAsp 127
DB 913 CAAGCAGAAATTT-----ATCTAAATTCACGCCGATGTAAATAATCTCTTCGCGCGAT 966
QY 128 AlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValProAlaIlePhe 147
DB 967 CAAGACCAATACATCAACATTTGAA-----ATTGCTCATCATGTCCTCAAAATTA--- 1017
QY 148 ThrArgGlyGluLeuAlaGlyArgTyThrValSerPheAspSerLeuLeuGluGlnPro 167
DB 1018 -----CAAAAATTGTAGCAATGAATCCACCATTCAA----- 1050
QY 168 GlyGlyIleGluAlaAsp-----AsnAlaPheAlaIaThrGlyPro 181
DB 1051 -----TTACACAGAAATCCTCTTCTTGCACTTGGAAGTTTATTAACAGGTTC 1104
QY 182 AspThr-----IleAla 185
DB 1105 AACGCCCAACAGATGAACCTAACCCAGCTTAACCAAAAACAAATTATCGATTATTT 1164
QY 186 LysPheLeuPheThrSerGlySerThrLysLeuProLysAlaValProThrThrglnArg 205
DB 1165 ACGATTATTATTAATCTCAGGCAACAGCGGAGACCTTAAGGTCTCATGTTAGATTACGCT 1224
QY 206 MetLeuCyAlaAsnGlnGlnMetLeuLeuGlnThrPheProValPheGlyGluGlnPro 225
DB 1225 AATCTCGCT-----CACCAATTAAGAAACACATGATCTTCACTTAATGTACAGATCAG 1278
QY 226 ProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGlySerHisAenIleGly 245
DB 1279 GATATTTCACCTTCTTTTTCACATTCCTGCATATTTTTCACAGGGCATGG-----GCGGCT 1335
QY 246 lIeValLeuTyraAsnGlyGlyThrTyrr--TyrlenuAspAspGlyLysProThrAlaGln 264
DB 1336 TATATTCTCATAGAGCGCAATATCTTGCTATTAGAAACACTAATCAATCAAGTCGGCTCA 1395
QY 265 GlyPheAlaGluThrLeuArgAenLeuSerGlnIleSerProThrAlaTyrlLeuThVal 284
DB 1396 GCT-----TTAACGGAATTCGCCCACTTAATGTGCGCGTA 1434
QY 285 ProLysGlyTrpGluGluLeuValGlyAla-----LeuGluArgAspSerThr 300
DB 1435 CCACGTTTTTACAAAATAATTTATGCTGCCGTATTGGATTAAGTTCAAAAAGCCCAAAA 1494
QY 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaIaGlyLeuSer 320
DB 1495 CTTCGCCAA-----ATTATGTTTCATTGGCAATTTCCGGGGA 1533
QY 321 GlnGlyIleTrpAsp--ArgLeuAspArgValAla----- 331
DB 1534 CAAAATAATTTGATCTACGTGCAATTAACAAAGCTATTCGTTCTTATTTGAAGAAACAA 1593
QY 332 -----GluGlnHisCySgIyGluArgIle 339
DB 1594 TTTCGTTTACAGATTAATTAATTAAGTCTCTCAAACTTCGTCAATTAATTTGGGGGACGATA 1653
QY 340 ArgMetMet----- 342
DB 1654 AAAATATATGCTTCGAGAGCAAAATTAAGAACCGGCTATTTGGGCTATTTTCCATGCT 1713
QY 343 -----AlaGlyLeuGlyMetThrGluThrAlaPro-----SerCyS 354
DB 1714 ATTGATATCAACATCAAAATTAAGCTATGTATGACGAACCAACTGCAACCGTTTCTTGC 1773
```

LOCATION: (47036) .. (47036)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (51334) .. (51334)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (51662) .. (51662)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (51786) .. (51786)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (51805) .. (51805)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (55369) .. (55369)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (65309) .. (65309)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (65313) .. (65313)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (80024) .. (80024)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (100091) .. (100091)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (102696) .. (102696)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (105121) .. (105121)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (107248) .. (107248)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (117136) .. (117136)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (119750) .. (119750)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (119924) .. (119924)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (120038) .. (120038)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (121344) .. (121344)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (122167) .. (122167)	

Pred. No.:	1,76e-16	Length:	1830121
Score:	283.00	Matches:	137
Percent Similarity:	38.44%	Conservative:	114
Best Local Similarity:	20.98%	Mismatches:	252
Query Match:	9.18%	Indels:	150
DB:	4	Gaps:	28

[illegible]

```

OY      301  LeuATGAGGAAAGPhenPhenAlaArgMetLeuLeuPhePhePheAlaAlaAlaGlyLeuSer 320
OY      1495  CTGGCCCAA-----ATTATGTTTCAATGGGCAATTTCCGTGGGA 1533
OY      321  GlnGlyIleTyrPaeP--ArgLeuAspTyrValAla----- 331
OY      1534  CAAAAATATTTTGAATCATCGTGGCAATACAAAGGTATTCGGTCTTATTAAGAAACAA 1593
OY      332  -----GluGlnIleGlyGlyGlnAlaGlyIle 339
OY      1594  TTGGCTTTAGCAGATTAATTAGTGCCTCCAAACCTTCCTCATTAATTGGGGGAGCATATA 1653
OY      340  ArgMetMet----- 342
OY      1654  AAAATGATGCTTGGCGAGGACCAAAATTAAGACCGGCTATTGGGCTATTTTCCAGTCT 1713
OY      343  -----AlaGlyLeuGlyMetThrGlnIthrAlaPro-----SerCys 354
OY      1714  ATTTGGTATCAACATCAATTAATGAGGTATGATGATGACAGAAACAACTGCACGACGTTTCTTGG 1773
OY      355  -----ThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaPro 372
OY      1774  TGGCATGATTTCCCATTTTAACCCAAATTTCA-----ATCGGACACATGATGCCA 1821
OY      373  GlyCysGlyValLeuLeuValProValAspGlyLeuGlyGlnAlaGlyArgPheHisGlyPro 392
OY      1822  AAAGCGGAAGGAAA-----ATTGGGAAAATAATGAATATCTTGTGGTGGCGGGA 1872
OY      393  HisValMetSerGlyTyrTyrPArgAlaProGluGlnAsnAlaGlnAlaPheAspGlyGln 412
OY      1873  ATGTGTGATGAAGGCTATNTACAAAGCCGAAACAAACCGCTCAAGCTTACACGGAAGAT 1932
OY      413  GlyTyrTyrCysSerGlyAspAlaIleLeuLeuAlaAspProAlaAspProGlnIleGly 432
OY      1933  GGTTCCTTAAACATCGCGCATGACAGAGAAATTT-----GACGAACAAAGGCAT 1980
OY      433  LeuMetPheAspGlyArgIleAlaGlnAspPheLeuLeuSerSerGlyValPheValSer 452
OY      1981  TTATTATTATTCGATCGGTATTAAGAAATTAATGAACCTTAACCGGCAATATATATGCGCA 2040
OY      453  ValGlyProLeuAspTyrThrArgAlaValLeuGlnGlyGlySerTyrValLeuAspValAla 472
OY      2041  -----CCACAATATATCGAAGCAAAATCGGTAAAGATAATTTATTCGAACAAATTCGG 2094
OY      473  ValAlaAlaProAspArgGlyCysLeuGlyLeuLeuValPheProArgLeuLeuAspCys 492
OY      2095  ATCATCGCTGATGCGRAAAATATGATATCGCGCTTATATGCGCT-----TGC 2142
OY      493  ---ArgAlaLeuSerGlyLeuGlyGlyGlnAlaSer-----AspAlaGly 506
OY      2143  TTGTATGATGTTTGAAGAGATACGCTTAAACAGCTCATTTAAATATCATATGACCGGTTTAGAA 2202
OY      507  ValLeuAlaSerGlyProValArgAlaThrPheAlaAspTyrLeuLysArgLeuAsnArg 526
OY      2203  CTACTTAAAAATTTGTGACATTTGCAAAATGTTTAGACATCGTATTAATTCGGGTGCAAAA 2262
OY      527  GluAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeuAspThrProProSer 546
OY      2263  GAA---TTGGCTCATTTGACAGCAAGTAAAAAAATTAACGTTACTTTCGCAAGCATTCAGC 2319
OY      547  IleAspLysGlyGlyValAlaThrAspLysGlySerIleAsnGlnAlaValLeuGlnThr 566
OY      2320  ATTAAATTTAGCGCAAAATTATCCACCAACTTAAATTAAGCTGAAAAAGCATTTTAGAACGT 2379
OY      567  ArgSerAlaLysValAspAlaLeuTyrArgGlyGlnAsp 579
OY      2380  TATCGCAAGCAAAATGAAGCATATTCATTCACACAGAA 2418

RESULT 8
US-10-329-960-1
; Sequence 1, Application US/10329960
; GENERAL INFORMATION:

```


QY	372	ProG1CYseAlValVlylsleu-----ValProValAaprgLylys	384
Db	33840	GCGGGCCTGCAGACACCAAGCTGTCTGAATCCGAGAGTCCAGTGCCCTCACACAGGGGAA	338938
QY	385	LeuGlUglYArphe-----HlsglyProHlsValwetsergLytyrTrparGalapro	402
Db	33900	ACCGTTGGGAGGCTTCAGTTCCTAGTCCGACGATGTTCGGCGCATTAATGACCG	339558
QY	403	GlUglnAsnAlaglnAlPheapRglUnglYTyrrCysserGIylerAlaleys	422
Db	33960	GAAGCCAGCACCCAGAGCGGTTCAGACGAGACGGTGTGACCTGACAGTAT--GTGGC	34016
QY	423	LeuAlAspProAlaaPProGlnLysglYLeuMetPheapRglyrAllealagluer	442
Db	34017	GTCTGTGACAGTGGCGGGAGTGCATCCG-----ATCGTGGGTCCGAGTCCGTTCAC	340678
QY	443	PheIylserSergerLyvalPhevalSerValGlyProleuarThraAlavalLeu	462
Db	34068	CATATCAATTGGGCGCGATACCAGATCCGACGACGAGAAATC-----CAGATGCCGCTG	341212
QY	463	GlUglYglYserTyrrValLeuaApvAl---ValValAlalaProasparGluCytsleu	481
Db	34122	CTAGAGATATCCCGATGTGCGGAGGTGGCAGTGTGCGGCTGCTAC--GAGGATTTG	341787
QY	482	GlyLeuLeuValPheproArgLeuLeuAprCyarAlaleusergLyseugLyysGlu	501
Db	34179	GGTCMAACGATCGTGTGATTCGTC-----GTAGCGCAGAG	342142
QY	502	AlAserAspAlaclValLeualaser-----Glu	511
Db	34215	GCATTAAATCCCAGTAGTGAATTAATCATATTGTTGCCAACAGCTTTCGATACAAGCG	342744
QY	512	ProValargAlatrPheAlaaSPTrbleuYArgLeuasnarGlnAlathrGlyasn	531
Db	34275	CCGGTGAAGTCCGCTTCGTGGACGGCGCTGCCGCG--AACCG--GATGGAAAAGTCT	343300
QY	532	AlAserArgilemetrPvalGlyLeuLeuAspThrProProserlleasplysGlygu	551
Db	34331	CAAAAAGCAACTGCTCTTCGAGGGCTGACGCGGCTACACAGCAGTAGTATTAAGCGGCG	343900
QY	552	ValThrAspIysglYserlleasnGlnArgAlavalLeuGlnTrp-----	566
Db	34391	ACGCACTCGTTGGACCGGACCGGGGCGGATGTGTGGTTGGTCGCGCGCCACTGT	344500
QY	567	ArgSerAlayValAspAlaleutyrr-----ArgGlyGlnaspGlnser	581
Db	34451	AGAGCTTCAAAGCCACGCGCAATGTGACCTAGTGTGTGTCAGAGGGGTTCGCGGTAGAGT	345101
QY	582	MeLeuArGaP	585
Db	34511	TTTTACGCGAC	34522

RESULT 6
US-09-557-884-1
Sequence 1, Application US/09557884
Patent No. 6506581

GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium

```

1      OPERATING SYSTEM: MS DOS V6.22
2      SOFTWARE: ASCII Text
3      CURRENT APPLICATION DATA:
4      APPLICATION NUMBER: US/09/557,884
5      FILING DATE: 25-Apr-2000
6      CLASSIFICATION: <Unknown>
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER: 08/476,102
9      FILING DATE: JUN-5-1995
10     ATTORNEY/AGENT INFORMATION:
11     NAME: Michelle S. Marks
12     REGISTRATION NUMBER: 41,971
13     REFERENCE/DOCKET NUMBER: PB186P3
14     TELECOMMUNICATION INFORMATION:
15     TELEPHONE: 301-309-8504
16     TELEFAX: 301-309-8439
17     INFORMATION FOR SEQ ID NO: 1:
18     SEQUENCE CHARACTERISTICS:
19     LENGTH: 1830121 base pairs
20     TYPE: nucleic acid
21     STRANDEDNESS: double
22     TOPOLOGY: linear
23     SEQUENCE DESCRIPTION: SEQ ID NO: 1:
24     US-09-557-884-1

```

Alignment Scores:	1.76e-16	Length:	1830121
Pred. No.:	283.00	Matches:	137
Score:	28.44%	Conservative:	114
Percent Similarity:	38.44%	Mismatch:	252
Best Local Similarity:	20.98%	Indels:	150
Query Match:	9.18%	Gaps:	28
DB:	4		
US-09-750-986D-30 (1-589) x US-09-557-884-1 (1-1830121)			
Qy	15	LeuGLuIaTgLeuGLuHIStrPaLbYThrArgProGLuInThrCyVaLaLaLaIaaRg	34
Db	646	GTTCATCGTATTCAACAACAAGCCAAACTCGTACAAATATGACCGCACTTCGTATMAA	705
Qy	35	AlAlaLaenGlyGluTPRArgArgIleSerYrLaLaGluMePheHlaaSnVaLaRgLa	54
Db	706	---GAACACGGCTTGCGCGAGATATCTCTTGAAAACTTTCAGAGCAACTCAATCAA	762
Qy	55	IleAlaInSerLeuLeuProTyrgLYeuSerLaGluArgProLeuLeuIleValSer	74
Db	763	CTTTCCTGAGCATTCCTTGCTCACAAATTTGACGACAAAGATMAAATGGCAATTTTGCC	822
Qy	75	GlyAaen-----AspLeuGLuHISLeuGLuLeuAlaPheGlyLa	87
Db	823	CATATATNGAACGTTGGACAAATCGTTGACATTCGGCACTTCAACAATTCGA-----	873
Qy	88	MetYrLaIaGlyIleProTyrgCyProValSerProLaIyrSerLeuLeuSerGlnAsp	107
Db	874	-----GCATCAACAGTCCCTATTTCAGCAACCAATACACCCGACG	912
Qy	108	LeuAlaLysLeuArgHISileValGlyLeuLeuGLuInProGlyLeuValPheAlaLaAsp	127
Db	913	CAACGACAGATTT-----ATCCTTAATCAGCGCCAGATGAAAAATCTCTTGCGCGCAT	966
Qy	128	AlAlaLaProPheGlnArgLaIleGLuThrIleLeuProAspAspValProLaIlePhe	147
Db	967	CMAAGCAATTCGATCAACAACATTCGAA-----ATTGCTCATCATTCGCCAATAATTA--	101
Qy	148	ThrArgGLyGluLeuAlaGlyArgArgThrValSerPheAspSerLeuLeuGLuInPro	167
Db	1018	-----CAAAAAATTTGTACAGTGAATATCCACCATTCAA-----	1055
Qy	168	GlyGLyIleGLuAlaLaAsp-----AsnAlaPheAlaLaIaThrGlyPro	181
Db	1051	-----TTACACAAAGATCCCTTCTTCTTGACCTTGGGAAAGTTTATTAAAAACAGTTCA	1100
Qy	182	AspThr-----IleAla	185

Db 18143 GCATTAGATCCGATGATTGTAATGTTGCCAACAGCTTCGATACATAAGCCG 18202
 QY 512 Pro-ValArgAlaTTPPhenAlaAspTTPLeuLysValGLeuAsnArgGluAlaThrGlyAs 531
 Db 18203 CCGCGGGAAGTCC-----GCTTCGTGGACGCGCTGCCGCAACGCGATGGGAAAGTCC 18256
 QY 531 nAlaSerArgIleMetTTPValGlyLeuLeuAspTTPProProSerIleAspLysGly 551
 Db 18257 TCAGAAAAGCATGTCCTTCGAGAGGCTGACTGCCCTACACAGATGATGATTAATGCGGG 18316
 QY 551 uValThrAspLysGlySerIleAsnArgAlaValLeuGlnTTP-----566
 Db 18317 CACGCCATGTTGCGACCGGCGGCGCGGATGTTGTTGGTGGCGCGGCACATG 18376
 QY 567 -ArgSerAlaLysValAspAlaLeuTyr-----ArgGlyGluAspGlnSe 581
 Db 18377 TAGAGCTTCAGAGCCACCGCATGTGACAGTATGTGTGTGAGGGGTCCGCTAGAG 18436
 QY 581 rMetLeuArgAsp 585
 Db 18437 TTTTTCAGCGGAC 18449

RESULT 5
 US-08-311-731A-138
 ; Sequence 138, Application US/08311731A
 ; Patent No. 6583266
 ; GENERAL INFORMATION:
 ; APPLICANT: SMITH, DOUGLAS
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
 ; NUMBER OF SEQUENCES: 411
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
 ; STREET: 600 ATLANTIC AVENUE
 ; CITY: BOSTON
 ; STATE: MASSACHUSETTS
 ; COUNTRY: USA
 ; ZIP: 02210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Releasee #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/311, 731A
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: GATES, EDWARD R.
 ; REGISTRATION NUMBER: 31,616
 ; REFERENCE/DOCKET NUMBER: C0044/7125
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/720-3500
 ; TELEFAX: 617/720-2441
 ; INFORMATION FOR SEQ ID NO: 138:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 35961 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: circular
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Mycobacterium leprae
 ; US-08-311-731A-138
 Alignment Scores:
 Pred. No.: 1,786-21 Length: 35961
 Score: 302.50 Matches: 150
 Percent Similarity: 37.50% Conservative: 69

Best Local Similarity: 25.68% Mismatches: 234
 Query Match: 9.82% Indels: 132
 Db: 4 Gaps: 27
 US-09-750-986D-30 (1-589) x US-08-311-731A-138 (1-35961)
 QY 44 SerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuProTyrGly 63
 Db 33039 TCGGTGCTGAGGCTGTCCGAGGTGCACCGCTGTGAGCGGTGTGCGCCGCGG----- 33092
 QY 64 LeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGlnIleuGlnLeu 83
 Db 33093 ---ACGCATTCGACCTGTGCGATTACT-----33119
 QY 84 AlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeu 103
 Db 33120 ---GGTTCCTGATGTCCTGCTGTGCGGTGTGCGGTGTGCGGTGTGCGGTGTGCGGTGTG 33173
 QY 104 LeuSerGlnAspLeuAlaLysLeuArgHisIleVal-----GlyLeu 117
 Db 33174 GTCGAACG-----CGCATATCTACCGCATTCGCGCGCGCGCTCG 33218
 QY 118 LeuGlnProGlyLeuValPheAlaAlaAspAlaIleProPheGlnArgAlaIleGlnThr 137
 Db 33219 CTGCGGTCCGGGGCC-----AGTCCGACCTCGCG-----33248
 QY 138 IleLeuProAspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThr 157
 Db 33249 ---CCGACCGGTGCGCGCATATC-----CCG 33272
 QY 158 ValSerPheAspSerLeuLeuGlnProGlyGlyIleGluAlaAspAsnAlaPheAla 177
 Db 33273 GTTCACTAGATCT-----CGCTCCGAAACCGCTATCC 33308
 QY 178 AlaThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuPro 197
 Db 33309 GACCGCTCACCTGATCCACCGGATGTGATGATCACCTCGGATCTACTCGGCGCCCT 33368
 QY 198 LysAlaValProThrThrGlnArgMetLeuCysAlaAsnGlnMetLeuGlnThr 217
 Db 33369 AAGGGGTACTGTGACCGCGGACGAGCATCGCGTGAATTTGACCGCTGCGCCAGGCC 33428
 QY 218 PheProValPheGlyGluGluProProValLeuValAspTTPLeuProTTPAsnHisThr 237
 Db 33429 TGGCAGTGCAGCGCCGATGAC-----GTTTGGTCAAGATGCGCGCTTTCACGTG 33482
 QY 238 PheGly-----GlySerHisAsnIleGlyIleValLeuTyrAsnGly 251
 Db 33483 CACGGCTGTACTCGTGTGCTTGATCGCTTCGATT-----33521
 QY 252 GlyThrTyrTTPLeuAspAspGlyLysProThrAlaGlnGlyPheAlaGlnThrLeuArg 271
 Db 33522 GGAATCGCTTGTACACACCGGAAACCCACATCCGACCGCTATGCTCAGGCT----- 33575
 QY 272 AsnLeuSerGluIleSerProThrAlaTyrLeuThrValIleProLysIleTTPGluGluLeu 291
 Db 33576 ---TGCTGTGACCGCGGGGATCTCTGTATTTCCGGGTTCCACCGATGCTGCGGCTG 33632
 QY 292 ValGlyAlaLeuGluArgAspSerThrLeuArgGluArgPheAlaArgMetLysLeu 311
 Db 33633 GTGCA-----GACGAGCGAGTT---GCCCGGCGCTTAAGACCGCGCGGCTA 33677
 QY 312 PhePhePheAlaAlaAlaGlyLeuSerGlnGlyIleTTPAspArgLeuAspArgValAla 331
 Db 33678 CTAGTCTGAGGATGCTCATTTGCGGTTCGCGTTCGATGATGATGATGATGATGATGATGATGAT 33737
 QY 332 GluGlnHisCysGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrGlnThrAla 351
 Db 33738 GGTATCGGCGCCATCGAACG-----TATGCGACGACGGAATCACG 33779
 QY 352 ProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAla 371
 Db 33780 ATCACCTGTGACGTTGCGGACGCGGTGACCGCGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 33839

TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 38494 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIUM LEPRAE
US-08-311-731A-24

Alignment Scores:
Pred. No.: 3,71e-22 Length: 38494
Score: 309.00 Matches: 149
Percent Similarity: 37.78% Conservative: 72
Best Local Similarity: 25.47% Mismatches: 231
Query Match: 10.03% Indels: 133
Gaps: 27
DB: 4

US-09-750-986d-30 (1-589) x US-08-311-731A-24 (1-38494)
QY 44 SerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGly 63
DB 16967 TCGGTGCTGAGCGGTGTGCGAGGTGCACGCTGTGTAGCGGTGCTGCCACGCCG----- 17020
QY 64 LeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeu 83
DB 17021 ---ACGCGATCGACCTGTGTCTAGCGATTACT----- 17047
QY 84 AlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeu 103
DB 17048 -----GTTGGCTGATTCGCGGTGTGCGCGTGTGCCCGCGCATTCGCGTGTG 17101
QY 104 LeuSerGlnAspLeuAlaIleValSerGlyHisIleVal-----GlyLeu 117
DB 17102 GTCGACGCG-----CGGCATATGCTCACCGATTCTGGGCGCGACGGCTTG 17146
QY 118 LeuGlnProGlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThr 137
DB 17147 CTGGGTCGCGGCGCC-----AGTGGCGACTCGCG----- 17176
QY 138 IleLeuProAspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgThr 157
DB 17177 -----CCGACCGGCTTGGCGCATATC-----CCG 17200
QY 158 ValSerPheAspSerLeuLeuGlnProGlyGlyIleGluAlaAspAsnAlaPheAla 177

DB 17201 GTTCACCTAATGCT-----CGTTCCTGAAACCGCATCT 17236
QY 178 AlaThrGlyProAspThrIleAlaIlePheLeuPheThrSerGlySerThrIleLeuPro 197
DB 17237 GAGCCGCACTGATGCGACCGGATGTGATCTACACCTCGGATACCTGGGCCCCCT 17296
QY 198 IysAlaValProThrGlnArgMetLeuCysAlaAsnGlnIleMetLeuLeuGlnThr 217
DB 17297 AAGGGGTACTGTTGAGCCGCGCGCATCGCTGATTTGACGCGCTTGGCCAGGCC 17356
QY 218 PheProValPheGlyGluProProValLeuValAspThrLeuProThrPheHisThr 237
DB 17357 TGGAGTGAACGCGCGATGAC-----GTTTGGTATAGATTTGCGGTTCACGTG 17410
QY 238 PheGly-----GlySerHisAsnIleGlyIleValLeuTyrAsnGly 251
DB 17411 CAGGCTTGTACTCGGTTTGCTTGATCGCTTCGGAAT----- 17449
QY 252 GlyThrTyrTyrLeuAspAspGlyIleProThrAlaGlnIlePheAlaGluThrLeuArg 271
DB 17450 GGAATCGCTTGTGACACCGGAAACCATCTCCGACCGCTATGCTCAGGCT----- 17503
QY 272 AsnLeuSerGluIleSerProThrAlaTyrLeuThrValProIleGlyIleProGluLeu 291
DB 17504 ---TGCTGTGAACCGCGGATCTGTATTTGCGCGTTCACCGATATGTCGCGCTG 17560
QY 292 ValGlyAlaLeuGluArgAspSerThrLeuArgGluArgPhePheAlaArgMetIleLeu 311
DB 17561 GTGCA-----GACGAGCGAGTT--GCCGGCGTTAAAGACCCCGCGGCTA 17605
QY 312 PhePhePheAlaAlaAlaGlyLeuSerGlnGlyIleThrAspArgLeuAspArgValAla 331
DB 17606 CTAGTGTCTGGAGTGCTGATTCGCGGTCCGGTATTTGATGCGGTAGCGACCTCACA 17665
QY 332 GluGlnHisCysGlyGluArgIleArgMetMetAlaGlyLeuGluMetThrGluThrAla 351
DB 17666 GGTTCATCGGCCCATCGAACCG-----TATGGACAGCAGGATCATCTG 17707
QY 352 ProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAla 371
DB 17708 ATACCTGTGCGACGTTGGCGGACCGGTGACCGCGCTGCGGTGGGATTTGCCCTG 17767
QY 372 ProGlyCysGlyValIleLeu-----ValProValAspGlyIle 384
DB 17768 GCGGCGGTGCACACGAGCTGTGATGATCCGAGTCCAGTGCCTTACGACGGGAA 17827
QY 385 LeuGluGlyArgPhe-----HisGlyProHisValMetSerGlyTyrThrArgAlaPro 402
DB 17828 ACCGTTGGGAGGCTTCAGGTTCTAGTCCGACGATGTTCCGCGCTATCTAAATCGACG 17887
QY 403 GluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleIle 422
DB 17888 GAAAGCAGACCGCAGGGGTTTGCAGAGAGGTTGGTACCTTACAGGTGAT--GTTGCG 17944
QY 423 LeuAlaAspProAlaAspProGlnIleGlyLeuMetPheAspGlyArgIleAlaGluAsp 442
DB 17945 GTCGTGACAGAGTGGCGGATGATCGC-----ATGCTGGTGTCCAGTGGGTTGAC 17995
QY 443 PheIleLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeu 462
DB 17996 CTAAATCAATTTGGCGGATATCCGATCGCGCGAGGAATC-----GAGATGCGCTG 18049
QY 463 GluGlyGlySerTyrValLeuAspVal-----ValValAlaAlaProAspArgIleCysLeu 481
DB 18050 CTAGGACATTCGGATGTGGGAGTGGGAGTGTGGCTGCTTAC--GAGGATTTG 18106
QY 482 GlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyIleGlu 501
DB 18107 GGTCAACGATCTTGCATTTCTC-----GTAGCGCGAGAG 18142
QY 502 AlaSerAspAlaGluValLeuAlaSer-----Glu 511

```

; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2393
; LENGTH: 1902
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2393

Alignment Scores:
Pred. No.: 5,756-175 Length: 1902
Score: 1652.50 Matches: 310
Percent Similarity: 70.57% Conservative: 100
Best Local Similarity: 53.36% Mismatches: 168
Query Match: 53.62% Indels: 3
DB: 4 Gaps: 3

US-09-750-986D-30 (1-589) x US-09-328-352-2393 (1-1902)

QY 5 GluAlaLeuLeuProPheProGlyArgGlyLeuGluLeuGluGluHisThr 24
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 GAGCAATTAAACCGTATCCACAAAACGACAGATGTTTATTCATTGTCACAAACC 132

QY 25 ArgProGluGlnThrCysValAlaAlaArgAlaAlaAnglyGluTrpArgArgLys 44
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 AAACCCGACCATATTTTGCAGCAAAACGCAATGCTCAAGCGCAATGGTCAAACGTAGT 252

QY 45 TyrAlaGluMetPheHisAsnValArgAlaAlaAlaGlnSerLeuLeuProTyrGlyLeu 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 TAGCAGAAAGTTTACAAACGCGATGCGACATGCTTCACCTTGATGCTCGTATTTA 312

QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuAla 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 AGCCAAAGAAAGACTCTGCTGCTATTTAAGTGTAATGATCTCGAACATTTTAAACATGCT 312

QY 85 PheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeu 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 ATGGCTGCATGCTGCGACGCGCTGCTTCTCGGCTATTTCCCTGCTACTCTGATTT 432

QY 105 SerGluAspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 433 TCCCAAGACTTTGGCAAACTCAACATGTTGAAAGTCTCAACCTGGTATGTTTAT 492

QY 125 AlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValPro 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 493 GCCAGCGATGACAGACTTTGGCCAAAGCATTCAGCATGATTAACGCTGATTTGAA 552

QY 145 AlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPheAspSerLeu 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 553 GTGGTGAACCAATAAAGAAATAGTGGCGATCAGATCGACAGTCTTTCAATCGCTGTA 612

QY 165 GluGlnProGlyIleGluAlaAspAsnAlaPheAlaIleThrGlyProAspThrIle 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 613 GATACACCAAGTTTCA--AATGTTCAAGATTTTATCAAAACCTTGATGAACCAAGATT 669

QY 185 AlaLysPheLeuPheThrSerGlySerThrLysLeuProLysAlaValProThrThrGln 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 670 GCCAAATTTCTGTTTACATCAGGTTCAACTAATTAACCTTAAGCTGACGACCAACAT 729

QY 205 ArgMetLeuCysAlaAngGlnIleMetLeuLeuGlnThrPheProValPheGlyGluGln 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 730 TTAAATGTTGTTGTTAATCAGCAAAATGTTATTCAGACTTCCCTGAGTTGAAGAAACA 789

QY 225 ProProValLeuValAspTrpLeuProTyrAsnHisIleThrPheGlyGlySerHisAsnIle 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 790 CCGCTGCTCTACTGACATGCGCTGCTCTGGCACCAACATTTGGCGGACATCAAAATGTC 849

QY 245 GlyIleValLeuLeuTyrAngGlyGlyThrTyrTyrLeuAspAspArgLysProThrAlaGln 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 850 GCGATGCACTCTATAACGCGGATGATGATTTACATGATATGCAAAACCGTTCAGACA 909

QY 265 GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrLeuThrVal 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 910 AATTTTGACAAACTATTCGTAATTCCAAGAAATTTCTCAACTGTTATTTAAATGTG 969

```

```

QY 285 ProLysGlyTyrGluGluLeuValGlyAlaLeuGluLysAspSerThrLeuArgGluArg 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 970 CCAAAAGGTTGGAGAAGACTCACCGAAGCCTTGAAAGAAATGAAAGATTAAGAGACCGC 1029

QY 305 PhePheAlaArgMetLysLeuPhePheAlaAlaAlaGlyLeuSerGlnGlyIleTrp 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1030 TTTTGGCCAAAGTTAAATTTATTTATTTCTTGGCGGTGCTGACATTTTCAGAAAGCGGCTGG 1089

QY 325 AspArgLeuAspArgValAlaGluGlnHisCysGlyGluArgGlyLeuArgMetMetAlaGly 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1090 AACGACCTCATTAATTTGCTCAGCAACATGGCGGAGAAAATCCGCATTAATGAGCGGA 1149

QY 345 LeuGlyMetThrGlnThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAla 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1150 TTGGCGATGACCAAACTGCTCTCTTGTGCTTTTACATGAGCCCAACCGGATAGCT 1209

QY 365 GlyTyrIleGlyLeuProAlaProGlyCysGluValLysLeuValProValAspGlyLys 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1210 GCGTTTATTTGTTATACCTGCTCGGGATGGAAATTAAGTATTCATGCTGATGTCACAA 1269

QY 385 LeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTrpArg--AlaProGlu 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1270 CTTGAGTTTTCGCTGCGCAAAACATCATGAAAGGCTATTGGCGCTTAAAGCGGAC 1329

QY 404 GluAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleLysLeu 423
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1330 CAACAAAGTACTATTTTGAATGAAAGCTTTTTCATACAGGCGATGCCGCTTCTTA 1389

QY 424 AlaAspProAlaAspProGluLysGlyLeuMetPheAspArgLysAlaGluAspPhe 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1390 GTGATGTCATTAATGATCCAAACCAAGGCTTAATGACGCAAGAAATGCGCAAGACTTT 1449

QY 444 LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1450 AAACCTAATACAGCACTTTTGTGAATGTCGACACATCGCAAAAGGCTTAATTCAA 1509

QY 464 GlyCysTyrValLeuAspValValAlaAlaProAspArgLysCysLeuGlyLeu 483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1510 GGTAAATTTACTCTCAACAGATGTTTGTATTTACTGCTCAAACTGAATGCTATTTGTTT 1569

QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGly--LysGluAla 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1570 CTGATTTTCCAAATTTAAGACGCTTGTCATTAATGACAGCTTTAAGTGGCGCAATAT 1629

QY 503 SerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuLys 522
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1630 TCTGCAAGAGATGACTGCAACACCTTAAGTCAACATGTTCCGCCAATTTTAAACG 1689

QY 523 ArgLeuAspArgLysAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAsp 542
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1690 ACCTTAATTAAGATGCGACTGCGACGCTCAAAATACGTTCAATGCTTTATTAATGACC 1749

QY 543 ThrProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGlnArgAla 562
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1750 GAGCCACCTCAGTAAATGCGCGGCGAAGTACCGAATGAAGCAACCTCAATCAAGACAGT 1809

QY 563 ValLeuGlnTrpArgSerAlaLysValAspAlaLeuTyrArgGlyGluAspGlnSerMet 582
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1810 ATTAACCAAGCTGTGCGCTTAAATGATGACCTTTATCAAAAGCACTGATTAACCG 1869

QY 583 Leu 583
Db 1870 CTG 1872

RESULT 4
US-08-311-731A-24
; Sequence 24, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR

```

NAME/KEY: CDS
LOCATION: (3146) .. (3997)
OTHER INFORMATION: Gene = "ORF1"
US-08-976-063E-1

Alignment Scores:

Pred. No.:	0	Length:	32679
Score:	3082.00	Matches:	589
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-09-750-986D-30 (1-589) x US-08-976-063E-1 (1-32679)

```
QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluHis 20
Db 21286 ATGGCTTCTCTCGAGGCGCTTCTTCCCTTCCCGGGTCGAATTCTTGAGCGTCTGAGCAT 21347
QY 21 TrpAlaIysThrArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTrp 40
Db 21348 TGGGCTAAGACCCGTCGAGAACAACTGCGTCTGCGCAGGGGGGCAATGGGGAATGCG 21407
QY 41 ArgArgIleSerTrpAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 21408 GCTGTATCAGCTACGCGGAATGTTCCACAACGTCGCGCATCGCACAGAGCTTGCTT 21467
QY 61 ProTrpGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 21468 CTTTACGAGCTATATCGAGACCGCTCGCTGCTTATCGTCTGGAATAAGACCTGGAACAT 21527
QY 81 LeuGlnLeuAlaPheGlyAlaMetTrpAlaGlyIleProTrpCysProValSerProAla 100
Db 21528 CTTGAGCTGGGATTGGGGCTATGATAGCGGGCATTCCTTATGCCGGTCTCCGCT 21587
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaIysLeuArgHisIleValGlyLeuLeuGlnPro 120
Db 21588 TATTTCATGCTGTCGCAAGATTGGCGAAGCTGCTCATCGTACATCGTACGCTTCTGCAACCG 21647
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
Db 21648 GAGCTGCTCTTGTGCGCATGACGACCTTTCAGAGCGCGCAATTGAGACCATTCGCGC 21707
QY 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgTrpValSerPhe 160
Db 21708 GAGCAGCTGCCCGCAATCTTCATCGAGGGAAATTGGCGGGCGCGCAGCGTAGTTT 21767
QY 161 AspSerLeuLeuGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
Db 21768 GACAGCTGCTGGAGCAGCCTGGTGGGATTGAGGCAATATATGCTTTGCGGCAACTGGC 21827
QY 181 ProAspThrIleAlaIysPheLeuPheThrSerGlySerThrIysLeuProValAlaVal 200
Db 21828 CCGGATTCGATTGCCAAGTCTGTGTCACTTGTGCTTACCAAACTGCTTAAGCGAGTG 21887
QY 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnIleMetLeuLeuGlnThrPheProVal 220
Db 21888 CCGAGCTACTCAGGAATGCTCTGCGCAATCAGCAATGCTTTCGAAACTTTCGCGGTT 21947
QY 221 PheGlyGluGluProProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGly 240
Db 21948 TTTTGTTGTAAGAGCGCGGTCGTGTGAGACTGTGTCGTTGCGTAACCACTTCGCGCGC 22007
QY 241 SerHisAsnIleGlyIleValIleuValAsnGlyGlyIleThrTrpTrpLeuAspAspGlyLys 260
Db 22008 AGCCACAACAATCGGCATCGTGTGTACACGCGCGCACTTGAACCTTGAAGAGGTAAA 22067
QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
Db 22068 CCAACGCCCAAGGTTCCGCGAGACGCTTCGAACTTGAGCCAAATCTCTCCACACTGCG 22127
QY 281 TyrLeuThrValProIysGlyTrpGluLeuValGlyAlaLeuGluArgAspSerThr 300
|||||
```

```
Db 22128 TACCTCACTGTGCCGAAGAGCTGGAGGAATTAAGTGGCTCCCTTGAGCGAGACGATACC 22187
QY 301 LeuArgGluArgPhePheAlaArgMetIysLeuPhePhePheAlaAlaAlaGlyLeuSer 320
Db 22188 CTCGGCGAAGCCTTCTTCGCTCCGATGAAGCTGTTCTTCTTCGCGCGGCTGGGTTCGCG 22247
QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGlyAlaArgIleArg 340
Db 22248 CAAGGAGATCTGGATCGTTTGAGCCGGGTGCTGTAACAGCACTGTGGTGAACGCGATTGCG 22307
QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360
Db 22308 ATGATGGCGGGTCTGGGCGATGACGAGACGCTCTTCCTGCGACCTTTTACCAACCGGACCG 22367
QY 361 LeuSerMetAlaGlyTrpIleGlyLeuProAlaProGlyCysGluValIysLeuValPro 380
Db 22368 CTGCGATGGCTGTTTACATTGGGCTGCGACCGCTGGCTGCGAGTCAACTGCTTCGCG 22427
QY 381 ValAspGlyIysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTrpArg 400
Db 22428 GTCCATGGGAAATTGGAAGGGCGTTTCCATGCTCCGACGTCATGAGCGGCTACTGGCGT 22487
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTrpTrpCysSerGlyAspAla 420
Db 22488 GCTCCTTAAACAAATGCCCCAAGCGTTGACGAGGAAAGCTATTACTGCTCGGTGATGCC 22547
QY 421 IleIysLeuAlaAspProAlaAspProGluIleGlyLeuMetPheAspGlyArgIleAla 440
Db 22548 ATCAAAATGGCAGATCCTGCGCATCTTCAGAAAGGCTTGATGTTGACGCTCGAATTCCT 22607
QY 441 GluAspPheIysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 22608 GAAAGCTTCAAGCTGCTCCTGAGGGGTATTGTGTCAGCGTTGGGCGCATTCGCGACCGGGCG 22667
QY 461 ValLeuGluGlyGlySerTrpValLeuAspValValAlaAlaProAspArgGluCys 480
Db 22668 GTTCTGGAAGCGGCTCTTACGTCCTGACGTAGTGGTGTGCTCTCGATCGGAATGC 22727
QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
Db 22728 CTTGATTTGCTCGTGTTCGCGCTTCTTCGACTGCGGCTTGTGCGGGCTTGGGAA 22787
QY 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
Db 22788 GAGCGCTCGAGCGCGCGAGGTCCTTCCAGTGAACCGGTTGGGCTGGTTTGCAGACTGG 22847
QY 521 LeuIysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeu 540
Db 22848 CTCMAAGACTCAATCGAAGAACTGGCAATGCCAGTGCATCATGTGGGTAGGGCTC 22907
QY 541 LeuAspThrProProSerIleAspIysGlyGluValThrAspIysGlySerIleAsnGln 560
Db 22908 CTCGATTCGCCCGCGCTGATTTGATGAAGGCGAGGCTCATCAAGGGCTCGATTAACAG 22967
QY 561 ArgAlaValLeuGlnTrpArgSerAlaIysValAspAlaLeuTrpArgGlyGluAspGln 580
Db 22968 CGGCTGTTTGAATGCGGTCGGAAGGTGATGCGCTGATCTGTGTGAAGATCAA 23027
QY 581 SerMetLeuArgAspGluAlaThrLeu 589
Db 23028 TCCATGCTGCGTGAACGAGGCCACACTG 23054
```

RESULT 3

US-09-328-352-2393
Sequence 2393, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04

US-09-750-986D-30 (1-589) x US-08-976-063E-29 (1-1770)

QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyValArgIleLeuGluValGluLeuHis 20
Db 1 ATCGCTTCTCTCCAGAGCGCTTCTCCCTCCCGGGTGAATTCCTGAGCGCTCTCGAGCAT 60
QY 21 TPRLAlaYethrArgProGluGluInThrCysValAlaAlaArgAlaAlaAngIyGluTrp 40
Db 61 TGGGCTAAAGACCCGTCAGAACAAACCTGCTGTGCTGCGAGCGCGCAAAATGGGAAATGG 120
QY 41 ArgArgIleSerTyraIaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 121 CGTCGATACAGCTACCGGAAATGTTCCAAACGTCGCCGCCCATCGCAGACAGCTTGCTT 180
QY 61 ProTyGlyLeuSerAlaGluValProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 181 CCTTACGGAATATCGGACAGCGCTCCGCTTATCGTCTGTGAAATGACCTGGAAACAT 240
QY 81 LeuGluLeuAlaPheGlyAlaMetTyraIaGlyIleProTyraCysProValSerProAla 100
Db 241 CTTCAGCTGGCATTTGGGCTATGTATGCGGCAATTCCTATATGCGCGGTCTCTGCT 300
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaIyLeuArgHisIleValGlyLeuLeuGlnPro 120
Db 301 TATTCACTGCTGCGCAAGATTGGCGAAGCTGCTCACAATCGTAGSTCTTCTGCAACCG 360
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluTrpIleLeuPro 140
Db 361 GGACTGCTCTTCTGCTCCGATGACGACCTTTCACAGCGCGCAATTGAGACCATTTGCGC 420
QY 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgAspGluSerPhe 160
Db 421 GAGAGAGTGGCCCAATCTTCACTCGAGGCAATGGCCGCGCGCGCAAGGTGAGTTT 480
QY 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
Db 481 GAAAGCTGCTGGAGACGCTGCTGGGATTGAGGCAATATATGCTTTGGCGCAACTGGC 540
QY 181 ProAspThrIleAlaIyPheLeuPheThrSerGlySerThrIyLeuPheProIyValAla 200
Db 541 CCCGATACGATCCCAAGTCTTGTTCATCTTGGCTCTACCAAACTGCTTAAGGCGGTG 600
QY 201 ProThrThrglnArgMetLeuCysAlaAsnGlnGluMetLeuGlnInThrPheProVal 220
Db 601 CGACACTACAGGAATGCTCTGCGCAATCAGCAAGATGCTTCTGCAAACTTTCGCGTT 660
QY 221 PheGlyGluGluProProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGly 240
Db 661 TTTCGTGAAGAGCGCGCGTCTGGTGGACTGGTTCGCGTGAACCAACACACTTCGCGCGC 720
QY 241 SerHisAsnIleGlyIleValLeuTyraAsnGlyGlyThrTyraIeAspAspArgIyLys 260
Db 721 AGCCCAACAATCCGCAATCGTGTGTACAAACGCGGCGCACTTAACCTTGAACGAGTAAA 780
QY 261 ProThrAlaGlnGlyPheAlaGluInThrLeuArgAsnLeuSerGluIleSerProThrAla 280
Db 781 CCAACGCGCCCAAGGTTGCGCGAAGCCTTGGCACTTGACGCAAACTCTCCACACTGCG 840
QY 281 TyrLeuThrValProIyGlyTyraGluLeuValGlyAlaLeuGluValAspSerThr 300
Db 841 TACCTCACTGCTGCGAAAGCTGCGAGGAATTAAGTGGTCCCTTGAGCGACAGACAGTACC 900
QY 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSer 320
Db 901 CTGCGGGAAGAGCTTCTTCCCTCCAGTAAGCTGTCTTCTTCCGCGGCGCTGGGTGTCTG 960
QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGlnGlnHisCysGlyGlyValArg 340
Db 961 CAAAGGATCTGGATCTTTGACCGGAGTCTGTAACAGACATGTGTGTGAGCGCATTCGC 1020
QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360

Db 1021 ATGATGCGGCGTCTGGGCACTGACGAGACTGCTCTTCTCCGACATTTTACACCGGACCG 1080
QY 361 LeuSerMetAlaGlyTyraIleGlyLeuProAlaProGlyCysGluValIySLeuValPro 380
Db 1081 CTGTGCATGGCTGGTTTACATTTGGGCTGCGAGCCCTGGCTGCGCAAGGTCAAGCTCGTCCG 1140
QY 381 ValAspGlyLysLeuGluGlyArgPheHisGlyProHisIleValMetSerGlyTyraTrpArg 400
Db 1141 GTTCATGGAAATTTGAAGGCGCTTTCATGTGTCGCGCATGAGCGGCTACTGGCGT 1200
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyraTyraCysSerGlyAspAla 420
Db 1201 GCTCTGAACAAATATCCCAAGCGTTGCAAGAGGAAGCTTATTAAGTCTCGGTGATGCC 1260
QY 421 IleIyLeuAlaAspProAlaAspProGluIyGlyLeuMetPheAspGlyValArgIleAla 440
Db 1261 ATCAAAATGGCAATCTTGGCATCTTCAGAAAGGTCTGATGTGTTACGGTCAAAATTCCT 1320
QY 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 1321 GAAAGCTTCAAGCTGCTCTCAAGGAGTATTTGTACGCTTGGGCAATTCGCCACGCGGCG 1380
QY 461 ValLeuGluGlyGlySerTyraIleAspValValAlaAlaProAspArgIyCys 480
Db 1381 GTTCTGAAAGCGCGCTTATACGCTCTGACGTAAGTGTGCTGCTCTGATCGTGAATGC 1440
QY 481 LeuGlyLeuLeuValIleProAlaGluLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
Db 1441 CTTCGATGTCTGCTGTTCCGCTCTTCGACTGCGCGCTTGTGCGGGCTAGAAATA 1500
QY 501 GluAlaSerAspAlaGluValIleAlaSerGluProValArgAlaIleTrpPheAlaAspTrp 520
Db 1501 GAGGCGTCCGAGCCCGAAGGTCTTGCACAGTACGAGCCGCTGGGCTGCTGCTGACTGG 1560
QY 521 LeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeu 540
Db 1561 CTCAAACGACTCAATGAGAAAGCACTGGCAATGCGACTGATGTGGTGAAGGCTC 1620
QY 541 LeuAspThrProProSerIleAspIyGlyGluValThrAspIyGlySerIleAsnGln 560
Db 1621 CTGGAATACGCGCGCTGATGTAAAGGCGAGGTCATGCAAGGAGCTCGATCAACAG 1680
QY 561 ArgAlaValLeuGlnTrpArgSerAlaIyValAspAlaIleuTyraArgIyGluAspGln 580
Db 1681 CGCGCTGTTTGCATAGCGCGTGGCGGAAGTTGATCGCTGTATGTGTGAAGATCAA 1740
QY 581 SerMetLeuArgAspGluAlaThrIleu 589
Db 1741 TCCATGCTGCTGACGAGGCCACACTG 1767

RESULT 2
US-08-976-063E-1
; Sequence 1, Application US/08976063E
; Patent No. 6524831
; GENERAL INFORMATION:
; APPLICANT: Steinhubel, Alexander
; APPLICANT: Pfelefer, Horst
; APPLICANT: Rademore, Jurgen
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
; TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; FILE REFERENCE: Bayer-9998-CAO
; CURRENT APPLICATION NUMBER: US/08/976, 063E
; CURRENT FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 32679
; TYPE: DNA
; ORGANISM: Pseudomonas sp.
; FEATURE:

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 7, 2005, 23:57:56 ; Search time 142 Seconds
(without alignments)
2948.272 Million cell updates/sec

Title: US-09-750-986D-30

Perfect score: 3082

Sequence: 1 MSLEALPPGRILRLERH.....KVDALYRGDSQMLRDEATL 589

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=x1p
-O=/cgn2.1/USPTO.spool_p/US09750986/runat.03022005.071744.18593/app.query.fasta.1.775
-DB=Issued_Patents_NA -OPMT=factap -SUFIX=nni -MIMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09750986 @CGN 1.1 105 @runat.03022005.071744.18593 -NCPU=6 -ICPU=3
-NO NMAP -LARGEOUTERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued Patents NA:*

1: /cgn2.6/prodata/1/ina/5A.COMB.seq:*
2: /cgn2.6/prodata/1/ina/5B.COMB.seq:*
3: /cgn2.6/prodata/1/ina/6A.COMB.seq:*
4: /cgn2.6/prodata/1/ina/6B.COMB.seq:*
5: /cgn2.6/prodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2.6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3082	100.0	1770	4	US-08-976-063E-29 Sequence 29, Appl
2	3082	100.0	32679	4	US-08-976-063E-1 Sequence 1, Appl
3	1652.5	53.6	1902	4	US-09-328-352-2393 Sequence 2393, Ap
4	309	10.0	38494	4	US-08-311-731A-24 Sequence 24, Appl
5	302.5	9.8	35961	4	US-08-311-731A-138 Sequence 138, Appl
6	283	9.2	1830121	4	US-09-557-884-1 Sequence 1, Appl
7	283	9.2	1830121	4	US-09-643-990A-1 Sequence 1, Appl
8	283	9.2	1830121	4	US-10-329-960-1 Sequence 1, Appl
9	266	8.6	1845	4	US-09-543-681A-2894 Sequence 2894, Ap
10	263.5	8.5	1639	4	US-09-396-154-10 Sequence 10, Appl
11	263.5	8.5	1639	4	US-09-396-154-43 Sequence 43, Appl
12	262.5	8.5	1639	4	US-09-396-154-11 Sequence 11, Appl

13	259	8.4	4403765	3	US-09-103-840A-2 Sequence 2, Appl
14	259	8.4	4411529	3	US-09-103-840A-1 Sequence 1, Appl
15	258	8.4	1639	4	US-09-396-154-42 Sequence 42, Appl
16	252.5	8.2	1639	4	US-09-396-154-6 Sequence 6, Appl
17	251	8.1	2416	4	US-08-592-126-61 Sequence 61, Appl
18	251	8.1	2416	4	US-09-168-595-61 Sequence 61, Appl
19	249.5	8.1	1639	4	US-09-396-154-8 Sequence 8, Appl
20	249.5	8.1	1666	4	US-09-602-628-1 Sequence 1, Appl
21	248.5	8.1	1639	4	US-09-396-154-9 Sequence 9, Appl
22	247.5	8.0	1639	4	US-09-396-154-7 Sequence 7, Appl
23	247.5	8.0	1666	4	US-09-602-628-3 Sequence 3, Appl
24	246.5	8.0	1650	1	US-08-354-240A-3 Sequence 3, Appl
25	246.5	8.0	1653	1	US-08-354-240A-5 Sequence 5, Appl
26	246.5	8.0	1685	4	US-09-602-628-5 Sequence 5, Appl
27	246.5	8.0	1686	4	US-09-602-628-7 Sequence 7, Appl
28	246.5	8.0	2337	4	US-10-066-130-20 Sequence 20, Appl
29	246.5	8.0	2674	4	US-10-066-130-19 Sequence 19, Appl
30	246.5	8.0	2771	4	US-10-066-130-18 Sequence 18, Appl
31	246.5	8.0	5860	4	US-10-066-130-17 Sequence 17, Appl
32	246.5	8.0	12614	4	US-09-577-424-1 Sequence 1, Appl
33	246.5	8.0	12844	4	US-09-743-194-19 Sequence 19, Appl
34	246.5	8.0	12850	4	US-09-743-194-22 Sequence 22, Appl
35	246.5	8.0	12851	4	US-09-743-194-21 Sequence 21, Appl
36	246.5	8.0	13073	4	US-09-743-194-20 Sequence 20, Appl
37	246.5	8.0	13654	4	US-09-743-194-28 Sequence 28, Appl
38	246.5	8.0	14194	4	US-09-577-424-3 Sequence 3, Appl
39	244.5	7.9	2782	4	US-10-140-002-205 Sequence 205, Appl
40	244	7.9	3634	4	US-09-814-915A-28 Sequence 28, Appl
41	244	7.9	8937	4	US-09-872-733A-8 Sequence 8, Appl
42	244	7.9	8937	4	US-09-872-733A-9 Sequence 9, Appl
43	243.5	7.9	2445	1	US-08-122-520C-8 Sequence 8, Appl
44	241.5	7.8	1722	1	US-08-718-425-1 Sequence 1, Appl
45	240.5	7.8	1650	1	US-08-354-240A-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-976-063E-29
; Sequence 29, Application US/08976063E
; Patent No. 6524831
; GENERAL INFORMATION:
; APPLICANT: Steinduchel, Alexander
; APPLICANT: Priefert, Horst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONFERYL
; TITLE OF INVENTION: ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; FILE REFERENCE: VANILLIC ACID AND THEIR USE
; CURRENT APPLICATION NUMBER: US/08/976,063E
; CURRENT FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: not required under old rule
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1767)
; OTHER INFORMATION: product = "Ferulaseure-Coa-Synthetase" / gene =
; OTHER INFORMATION: "fcs"

Alignment Scores:

Pred. No.: 0
Score: 3082.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Length: 1770
Matches: 589
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
DB:

This Page Blank (uspto)

[illegible]

OY	236	GIUARGAPSErThrLeuargGIUARGPhephealAargMetLysLeuPhepheheAla	315
Db	1099	-----ACAAGTGCCTGTAATAATCATCTGGGGTTGTGAAAAAAGGCTTTTCGAAATTGSC	1152
OY	316	AlaIaGIuLeuSerGIuNGlyLe-----TTPAspArg	326
Db	1153	TATACTCAAGAGACCAAGCAGATCATTAATGGCCGACTCCTTCGCATTTTGGGACAAAG	1212
OY	327	Leu-----AspArgValaIaGIuGlnHisCysGIuArgGIIeaArgMetMetAla---	343
Db	1213	CTGGCTTCACACAAAATAAAAGAAAG---CTTGATGACCGGCTTCGGTTTATGGGGTCT	1268
OY	343	-----	343

Db	1270	GGTGGCTTCCTCTTGTACACGTAGTCATGAGATTTCCTTGAGAAATATGCTTTGGATGTTGC	1329
Oy	344	-----GlyLeuGlyMetThrGluThrAlaProSerCysThrPheThr-----	358
Db	1330	GTGCGTGAAGGATATGATATACCGAGACT-----TCTTGTCATTAAGCGCTATGAT	1383
Oy	359	---GlyProLeuSerMetAlaGlyTyrTrilegGlyLeuProAlaProGlyCysGlyValIlys	377
Db	1384	GATGCTGAC--AAATTATCTGGCCACGTGGATCCCTTAATCCAGCTTGAGAGTAAG	1440
Oy	378	LeuValProValAsp-----GlyIlys	384
Db	1441	CTTGIGGATGTTCCCGAATAATTAACATTCAGATCGGAAGATCAACATCCACGTGGTGA	1500

[illegible]

QY	458	-----ThaArgAlaValLeuGluGlySerTyr---	467
Db	1702	AATGTTATACCAATGTAGATTCGTTTCGAGGTTTCATTACGGTAGATGCTTCAAT	1761
QY	468	-----ValLeuAspValValAlaAlaIleProAsp-----	477
Db	1762	TCCTCTCTAGAGCTATAGTTTCAGTCGACCCCGAAGTTATGAAGAATTGGCGCTGCATCA	1821
QY	478	-----ArgGluCysLeuGlyLeuIleuVal---PheProArgLeuLeuAsp---	491
Db	1822	GAAAGCATCAAGTATGAGCACTTGGACAGCTCTGTAAACATCCAAAGTCCGAAAGACT	1881
QY	492	---CysArgAlaLeuSerGlyLeuGlyLysGluAla-----Ser	503

[illegible]

Search completed: February 8, 2005, 00:16:11
Job time : 977 secs

KW Genome; bacterium; Haemophilus influenzae; computer readable medium; expression modulating fragment; regulation; gene expression; vector; KW organism; open reading frame; ORF; de.

XX Haemophilus influenzae.

XX PN W09633276-A1.

XX PD 24-OCT-1996.

XX PF 22-APR-1996; 96WO-US005320.

XX PR 21-APR-1995; 95US-00426787.

XX PR 07-JUN-1995; 95US-00476102.

XX PR 07-JUN-1995; 95US-00487429.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (UYJO) UNITV JOHNS HOPKINS.

XX PI Fleischmann RD, Adams MD, White O, Smith HO, Venter JC; WPI; 1996-465782/48.

XX DR Haemophilus influenzae Rd genome recorded on computer readable medium - useful for identifying commercially important nucleic acid fragments by homology searching.

XX PT

XX PS Claim 1; Page 77.2-77.1091; 1291pp; English.

XX CC This sequence represents the complete genome sequence of the bacterium Haemophilus influenzae strain Rd. The invention relates to a computer readable medium (CRM) having recorded upon it the complete H. influenzae nucleotide sequence (I), a representative fragment of (I) or a nucleotide sequence at least 99% identical to (I). By providing the full-length genomic sequence in a computer readable form, it is possible to identify commercially important nucleic acid fragments and expression modulating fragments (EMFs) of the Haemophilus genome. The EMFs can be used to regulate the expression of a nucleic acid molecule. Vectors and altered organisms comprising the predicted ORFs can be used to produce any of the polypeptide fragments of the H. influenzae Rd genome

XX CC

SQ Sequence 1830121 BP; 567399A; 350615C; 347389G; 564036T; 0U; 6820cher;

Alignment Scores:

Pred. No.:	3.17e-13	Length:	110000
Score:	283.00	Matches:	137
Percent Similarity:	38.44%	Conservative:	114
Best Local Similarity:	20.98%	Mismatches:	252
Query Match:	9.18%	Indels:	150
		Gaps:	28

US-09-750-986d-30 (1-589) x AAT42063_00 (1-110000)

QY 15 LeuGluArgLeuGluHisTrpAlaLysThrArgProGluGlnThrCysValAlaAlaArg 34

DB 646 GTTCATCGTATTCACAAACAAAGCCAAAGCTCGTAAATAATGACCGCACTTCGCTTAA 705

QY 35 AlaAlaAsnGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 54

DB 706 ---GAACACGCGCTTGCGCAAGATATCTCTTGAAAACTTCAAGCAACCAATCA 762

QY 55 IleAlaGlnSerLeuProGlyLysLeuSerAlaGluArgProLeuLysLysLysLysLys 74

DB 763 CTTTCTCGAGCATTTGCTTCTCTCAATATTTGACGTACAGATATAAATCGCATTTTGGCC 822

QY 75 GlyAsn-----AspLeuGluHisLeuGluLeuAlaPheGlyAla 87

DB 823 CATTAATATGGAAGCTTGGACATCGTTGACATTCGACCTTACCAATTC 873

QY 88 MetTrpAlaGlyLysProGlyCysProValSerProAlaTyrSerLeuLeuSerGlnAsp 107

DB 874 -----GCAATCACAGTGCCTATTTCACCAACCAATACAGCCAG 912

QY 108 LeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValAlaPheAlaAlaAsp 127

DB 913 CAAGCAGAAATTT-----ATCCTAAATCAGCCGCAATGTAATAAATTCCTTCTGGCGCAT 966

QY 128 AlaAlaProPheGlnAlaIleGluThrIleLeuProAspAspValProAlaIlePhe 147

DB 967 CAAGAGCAATACGATCAAAACATTCGAA-----ATTCTCATTCATTCGCAAAATTA--- 1017

QY 148 ThrArgGlyLeuLeuAlaGlyArgThrValSerPheAspSerLeuLeuGlnPro 167

DB 1018 -----CAAAAATTGTACCAATGAAATCCACCATTCAA----- 1050

QY 168 GlyGlyIleGluAlaAsp-----AsnAlaPheAlaAlaThrGlyPro 181

DB 1051 -----TTACACAAAGATCTCTTCTTTCGACTTGGAAAGTTTATTAAACAGTTCA 1104

QY 182 AspThr-----IleAla 185

DB 1105 AACGCCCAACAGATGAATTAACCAACGCTTAACCAACCAAAACAAATTAATCGATTTATTT 1164

QY 186 LysPheLeuPheThrSerGlySerThrLysLeuProGlyValAlaProThrThrGlnArg 205

DB 1165 ACGATTATTTATACCTCAGGCAACGAGAGCTTAAGGTCATGATTAATTAAGCTT 1224

QY 206 MetLeuCysAlaAsnGlnGlnMetLeuLeuGlnThrPheProValPheGlyGluPro 225

DB 1225 AATCTGCT-----CACCAATTGAACACATCATCTTCACCTTAATGACAGATCAG 1278

QY 226 ProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGlySerHisAsnIleGly 245

DB 1279 GATATTTCATCTTCTTTTTCATCTCTCAATTTTTCACGCGCATGG---GCGGCT 1335

QY 246 IleValLeuLysAsnGlyGlyThrTrp---TyrLeuAspAspGlyLysProThrAlaGln 264

DB 1336 TATATTCTTATAGAGCGCAATATCTTGCTATTTTAAAGACCTAATCAAGTGGCTCA 1395

QY 265 GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrLeuThrVal 284

DB 1396 GCT-----TTACCGAAATTCGCCCACTTAAATGTGGCGCTA 1434

QY 285 ProLysGlyTrpGluLeuValGlyAla-----LeuGluArgAspSerThr 300

DB 1435 CCACGTTTTCAGAAAATTTATGTCGCGTATGGAATGAAGTCAAAAAGCCCAAAA 1494

QY 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaGlyLeuSer 320

DB 1495 CTTGCGCAA-----ATTATGTTTATTTGGCAATTCCTCGTGA 1533

QY 321 GlnGlyIleTrpAsp---ArgLeuAspArgValAla----- 331

DB 1534 CAATAATATTTTATCTACGTGGCAATACAAAGCTATTCCTTCTTATTTGAAGAAACAA 1593

QY 332 -----GlnGlnHisCysGlyGluArgIle 339

DB 1594 TTTCCTTAGCAGATTAATTAATGCTCTCAAAACTTCGTCAATTAATTTGGGGGACGATA 1653

QY 340 ArgMetLeu----- 342

DB 1654 AAAATGATGCTTCGAGAGCAACAAATTAAGAACCGGCTATTTGGCTATTTCCATGCT 1713

QY 343 -----AlaGlyLeuGlyMetThrGluThrAlaPro-----SerCys 354

DB 1714 ATTGATATCAATCAATTAATTAAGCTATGATACGAAACAACTGCAACCGTTTCTGCG 1773

QY 355 -----ThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaPro 372

DB 1774 TGGCAGATTTTCCAAATTAACCAATTC-----ATCGGACACATGATGCGCA 1821

QY 373 GlyCysGlyValLysLeuValProValAspGlyLysLeuGluGlyArgPheHisGlyPro 392

DB 1822 AAAGCGGAAGTGAAA-----ATTGGGGAATAAATGAATTCCTTGGCTGCGCGGA 1872

QY 393 HisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGlu 412


```

Db      1315 TGCTTTGGATGTTGCGTGCCTGAGGATGATGATGACCGAGACT-----TCTTGTGTC 1368
Qy      356 PhetThr-----GlyProLeuSerMetLeuArgLysLeuProAlaPro 372
          |||
          |||
          |||
Db      1369 ATAAAGTCTCTATGATGATGATGAC---AATTATTCGCGCATGTCGTTCCCTTAATCCA 1425
Qy      373 GlyCysGluValIleuValProValAsp----- 382
          |||
          |||
          |||
Db      1426 GCTTGCAGGTAATACTTGATGATGTTCCCGAATAATGATTAACATCAGACATCAACCA 1485
Qy      383 -----GlyIleuValGluValArgPheHisGlyProHisValMetSerGlyTyrTrp 399
          |||
          |||
          |||
Db      1486 TACCCAGCTGCTGTAATCTGTGTAAGA-----GACCAATCACTTCCAAAGGCTACTAC 1539
Qy      400 ArgAlaProGluIleuValAlaGlnAlaPheAspGluGluGlyTyrCysSerGlyAsp 419
          |||
          |||
          |||
Db      1540 AAGATATAAGAACAAACGAGAGAAATTCTGATGAGATGGCTGCTACACACAGAGAT 1599
Qy      420 AlaIleuValLeuAlaAspProAlaAspProGluIleuMetPheAspGly----- 437
          |||
          |||
          |||
Db      1600 ATC-----GGTTGTGCTTACCTGCTGTCGG 1626
Qy      438 -----ArgIleAlaGluAspPheIleuSerSerGlyValPheValSer 452
          |||
          |||
          |||
Db      1627 CTCAGATCATAGACAGAGAAAGAACATATTAAAGTTGGCGCAAGAGATATATAGCA 1686
Qy      453 ValGlyProLeuArg-----ThrArgAlaValLeuGlu 463
          |||
          |||
          |||
Db      1687 CCAGAGAAAGATCGAAATAATGTTATACAAATGATGATTCGTTTGGCAGTGTTCATTAC 1746
Qy      464 GlyGlySerTyr-----ValIleuAspValValAlaAlaProAsp----- 477
          |||
          |||
          |||
Db      1747 GGTGATAGCTTCAATTCCTCTAGTAGCTATAGTTCAAGTCGACCCCGAAGTTATGAA 1806
Qy      478 -----ArgGluCysLeuGluIleuVal---PhePro 487
          |||
          |||
          |||
Db      1807 GATTTGGCTGCATCAGAGAGCATCAATGATGAGCATCTAGACAGCTCTGTAAGATCCA 1866
Qy      488 ArgLeuLeuAsp-----CysArgAlaLeuSerGlyLeuGlyValAla----- 502
          |||
          |||
          |||
Db      1867 AGAGTGCAGAAAGACTGTTCTTCTGATGATGATGATCACTTGAGAAAGAGCTCACTTGACA 1926
Qy      503 -----SerAspAlaGluValIleuAlaSerGluProVal----- 513
          |||
          |||
          |||
Db      1927 GGGTTGAGTTGCAAGGCTGTGACTTTGGTGCAGAACCATTTGAGACCTTGAGAGATGGA 1986
Qy      514 -----ArgAlaTrpPheAlaAspTrp 520
          |||
          |||
          |||
Db      1987 CTTCTCACACCAATTCAGATTAAGAGACCTTCAGCAAAAGCCTTCTTGACAGAGCA 2046
Qy      521 LeuLysArgLeuAsnArgGlu 527
          |||
          |||
          |||
Db      2047 ATTAGCAAAATGATGCGGAA 2067
          |||
          |||
          |||
RESULT 13
ADN72468 standard; cDNA; 2103 BP.
AC      ADN72468;
XX      15-JUL-2004 (first entry)
DE      Thale cress cDNA upregulated in E2Fa/Dpa expressing plants Seqid 363.
KW      gene; ss; plant; transgenic; E2Fa/Dpa transcription factor;
KW      growth regulator; animal feed product; thale cress;
KW      cell wall biosynthesis; nitrogen metabolism; carbon metabolism.
OS      Arabidopsis thaliana.
XX      MO2004035798-A2.
XX

```

```

PD      29-APR-2004.
XX
XX      20-OCT-2003; 2003MO-BP011658.
PF
XX      18-OCT-2002; 2002EP-00079408.
PR
XX      (CROP-) CROPEDESIGN NV.
PA
XX      Inze D, De Veylder L, Vlieghe K;
XX      WPI: 2004-348466/32.
DR      P-PSDB; ADN72469.
XX
XX      Altering plant characteristics, useful for producing plants for enzyme or
PT      pharmaceutical production comprises modifying in a plant, expression of
PT      one or more nucleic acids and/or modifying level or activity of one or
PT      more proteins.
XX
XX      Claim 1; SEQ ID NO 363; 134pp; English.
PS
XX
XX      This invention relates to a novel method for altering one or more plant
CC      characteristics. Specifically, it refers to identifying genes that are up
CC      - or down-regulated in transgenic plants overexpressing the heterodimeric
CC      E2Fa/Dpa transcription factor of Arabidopsis and using these sequences to
CC      alter plant characteristics accordingly. The present invention describes
CC      generating transgenic plants for the production of growth regulators,
CC      enzymes, therapeutics, pharmaceuticals and animal feed products, where
CC      the altered plant characteristics are selected from increased yield or
CC      biomass, enhanced survival capacity, stress tolerance, plant architecture
CC      or physiology, altered endoreduplication, biochemistry, signal
CC      transduction, storage lipid mobilisation and/or altered photosynthesis,
CC      each relative to the corresponding wild type plants. Accordingly, these
CC      sequences can also be useful as positive or negative selectable markers
CC      during transformation of cells or tissues. The identified genes play a
CC      role in a variety of biological processes such as DNA replication, cell
CC      wall biosynthesis, nitrogen and/or carbon metabolism or they function as
CC      transcription factors. This polynucleotide sequence is thale cress cDNA
CC      upregulated 1.3 fold or more in plants overexpressing the E2Fa/Dpa
CC      transcription factor, given in an exemplification of the invention.
XX
SQ      Sequence 2103 BP; 589 A; 418 C; 496 G; 600 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1,49e-15 Length: 2103
Score: 283.00 Matches: 150
Percent Similarity: 37.03% Conservative: 97
Best Local Similarity: 22.49% Mismatches: 214
Query Match: 9.18% Indels: 206
DB: 12 Gaps: 35
US-09-750-986D-30 (1-589) x ADN72468 (1-2103)
Qy      9 ProPheProGlyArgIleuGluArgLeuGluHisTrpAlaLysThrArgProGluGln 28
          |||
          |||
          |||
Db      241 CCGTAATCGGAGCTTTCATCATGACAATTTGTGATGCTGTGTAACATGCTGAAC 300
Qy      29 ThrCysValAlaAlaArgAlaAlaAsn-----GlyGluTrpArgArgIleSer 44
          |||
          |||
          |||
Db      301 AAGTATCTTGTCACAGAGTTCGCTCCGATGCAACCATTTGAGAGATTTATGATGATACA 360
Qy      45 TyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeu 64
          |||
          |||
          |||
Db      361 TATGAGAACACAGGCTGTGAGCGCAAGCATTTGTTGCTGTTTCAAGAGTT 420
Qy      65 Ser-----AlaGluArgProLeuLeuIleValSerGly 75
          |||
          |||
          |||
Db      421 AACCAAGAGATTGCGTTGACTCTATTATTATTAACGACCAAGTGGTTGTT----- 474
Qy      76 AsnAspLeuGluHisLeuGluIleuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCys 95
          |||
          |||
          |||
Db      475 -----GTGATCATGCTTGTGACGACATATTCATTCTCTCTGTT----- 513
Qy      96 ProValSerProAlaTyrSerLeuLeuSerGlnAspLeuAlaLys-----LeuArgHis 113
          |||
          |||
          |||

```


WP ACF67367_54 5400001 5510000
WP ACF67367_55 5500001 5610000
WP ACF67367_56 5600001 5648994

Alignment Scores:
Pred. No.: 7.33e-14 Length: 110000
Score: 290.00 Matches: 146
Percent Similarity: 40.13% Conservatave: 110
Best Local Similarity: 22.88% Mismatches: 264
Query Match: 9.41% Indels: 118
DB: 10 Gaps: 26

US-09-750-986d-30 (1-589) x ACF67367_48 (1-110000)

QY 11 ProGlyArgIleuGluArgIleuGluHieTrpAlaIysThrArgProGluGlnThrCys 30
Db 59870 CCTTATACCTTGATTAACCGGTTTGCAACAAATMAACACACCCAGAGAAA----- 59923
QY 31 ValAlaAlaArgAlaAlaAsnGlyGluTrpArgArgIleSerTyraIaGluMet---Phe 49
Db 59924 ATTGCTTTTGC-----CAATGGCTCCGGCTGAACAACCTGAACTTAACCTGG 59971
QY 50 HieAsnVal-----ArgAlaIleAlaGlnSerIleuProTyryGlyLeuSer 65
Db 59972 CATGATATTGGCAAAAAGACCGACGACTTATAGCCGGCGCTTGTGGAGATTGGTGTGAT 60031
QY 66 AlaGluArgProIleuIleValSerGlyAsn-----AspLeu 78
Db 60032 GTGCAGAAAAAATTTGGCTCTTTGGCACTAACAGCATGGCGTGTCAATTGGCAGATATC 60091
QY 79 GluHieIleuGlnIleuAlaPheAlaIaAspAlaAlaProPheGlnArgAlaIleGlu---Thr 137
Db 60092 GCTGTATTACAACTTGG-----GGGTCACATGTC 60121
QY 99 ProAlaTySerIleuIleuSerGlnAspLeuAlaIysIleuArgHieIleValGlyLeuLeu 118
Db 60122 CCTCTGATCCAAACCACTAGTCAGGAAACAGGCTGGTTTC-----ATACTTAATGATGCT 60175
QY 119 GluProGlyLeuValPheAlaIaAspAlaAlaProPheGlnArgAlaIleGlu---Thr 137
Db 60176 GGGTGGCGATTTTGTGTGGTGGGAGGCGAGATGATGCTGCTATTAACCTTGACA 60235
QY 138 IleLeuProAspValProAlaIlePheThrArgIlyGluIleuAlaIaArgArgThr 157
Db 60236 GAATGTTGCCCTCAGTTAAGTCACATTATGTCTTAGATGATCAGTCGATTTGCCGGAT 60295
QY 158 ValSerPheAspSerIleuLeu-----GluGlnProGlyIle 170
Db 60296 TGTTCATTAGCTCAGCATTTATCTGATTCATGGTCGATGATTAATTAACCTGTCACTCC 60355
QY 171 GluAlaAspAsnAlaPheAlaIaIaThrGlyProAspThrIleAlaIaIysPheIleuPheThr 190
Db 60356 GAGTTGAAAACCTGATTTGCTTAACACCAATCTGGATATCTTTTACACTTATTTATACG 60415
QY 191 SerGlySerThrIlyLeuProIlySalaValProThrThrGlnArg-----Met 206
Db 60416 TCGGTATCCACGGGGTAGCCTTAAGGGTCATGCTTATTAACCGCAATTAGCCTATCAG 60475
QY 207 LeuCysAlaAsnGlnMetIleuLeuGlnThrPheProValPheGlyGluGluProPro 226
Db 60476 TTGTATTACATGATCATATCTCTGGCATTTAACT-----GATCAGAC 60517
QY 227 ValLeuValAspTrpLeuProTrpAsnHieThrPheGlyGlySerHieAsnIleGlyIle 246
Db 60518 ATTTCCTTGGTTTTCGCGTTATCCCATGCTTTTACGGCTGATGAGATTTCTATGTC 60577
QY 247 ValLeuTyAsnGlyIlyThrTyryLeuAspAspIlyIysProThrAlaGlnGlyPhe 266
Db 60578 ATGCATGTAGGGGACAAATGTTTATCTACGTACACTAATTTTGTCTCCAAAGCATG 60637
QY 267 AlaGluThrIleuArgIleuSerGluIleSerProThrAlaTyryIleuThrValProIys 286
Db 60638 GCG-----GATGCTGCCCGACAGTGAATGATGTGACAGTCCGAGA 60676

QY 287 GlyTrpGluIleuValGlyAlaLeu---GluArgAspSerThrLeuArgGluArgPhe 305
Db 60677 TTTTACGAAAAAATTTTACTGCGGCTTTTGAAAAGCTTCT-----AAGCTCCT 60727
QY 306 PheAlaArgMetIlyLeuPhePheAla---AlaAlaGlyLeuSerGlnGlyIle--- 323
Db 60728 TGGCATCGAAGAAAACCTTTTAATTGGCGCATTAATAGTGCGTACGCTGATTTTACGT 60787
QY 324 -----TrpAsp---ArgIleuAspArgValAlaGluGln 333
Db 60788 AAGTTCAGAGTAAAAAACCTGCCCTTGGTCAACAGCAATGATGCACTGGCTGATPAA 60847
QY 334 -----HieCysGlyGluArgIleArgMetAla----- 343
Db 60848 TTGGCTTGAACAAATTACGGAATGTGTGGGGGACAGATTCGGTTTTTACCTGCTCT 60907
QY 343 ----- 343
Db 60908 GGTCACGCTGGACGAAACGATTATCTGTCTTCTGCTGTGGGATTAATATCAAA 60967
QY 344 ---GlyLeuGlyMetThrGluThrAlaPro-----SerCysThrPheThrThrGlyPro 360
Db 60968 TATGGCTATGGCATGACAGAGACTTGTGCCACGGTTTCTGTGGGAAGCAAGT--- 61024
QY 361 LeuSerMetAlaGlyTyryIleGlyLeuProAlaProGlyCysGluValIysLeuValPro 380
Db 61025 ---TATGACTTGGTCTTATGTGAACCATGCCCTGGCATTAACGCTCCGATCCGGTAA 61081
QY 381 ValAspGlyIysLeuGluGlyArgPheHieGlyProHieValMetSerGlyTyryTrpArg 400
Db 61082 GAAGAT-----GAGATACAAAGTCCGTGCCCAATTTGATGAAGAAGCTATTCAAT 61132
QY 401 AlaProGluIleAsnAlaIaPheAspGluGluGlyTyryTyryCysSerGlyAspAla 420
Db 61133 AATCAGCAAAAAACCAACAGGATTTTACTCAAGACGGTGGTGCCTACAGGATGAGCC 61192
QY 421 IleIysLeuAlaAspProAlaAspProGlnIlyGlyLeuMetPheAspGlyArgIleAla 440
Db 61193 GGTAAAGTTG-----GATGAATAATGTAATTTGTCTACTTACGAAACGCTGAAA 61240
QY 441 GluAspPheIysLeuSerSerGlyValPheValSerValGlyProIleuArgThrArgAla 460
Db 61241 GATTTGATGAACATCGAATGCAATATATATGCCC-----CCTCAATGATTAAGAGC 61294
QY 461 ValLeuGluGlyIlySerTyryValIleuAspValValAlaAlaIaProAspArgGluCys 480
Db 61295 ATTTTGGACAAAGATCGTTTATTGAAACATGTGCTATTTTCTGTATGCTCGTAATTT 61354
QY 481 LeuGlyLeuIleuValPhePro-----ArgLeuLeuAspCysArgAlaIleuSerGlyLeu 498
Db 61355 GTATCGGCTTTTATTCGCTTGTCTCATGAAGCATTTGGAAAGATGATGCAAGTCAATTAAT 61414
QY 499 GlyIysGluAlaSerAspAlaGluValIleuAlaSerGluProValArgAlaTrpPheAla 518
Db 61415 CTCAAATATCATGACCGGCTGTGAATTAAGTCGTCATAGTCAGATTAATCATATGTTGAG 61474
QY 519 AspTrpLeuIlyArgIleuAsnArgGluAlaIaThrGlyAsnAlaSerArgIleMetTrpVal 538
Db 61475 CAACGCTTGAAAATAATGCAGAAAAC---TTTGTCTCACTTCGATCAGGTAAGCGTTTC 61531
QY 539 GlyLeuIleuAspThrProProSerIleAspIlyGlyIleValThrAspIlySerIle 558
Db 61532 ATATTTACTGCCAAAACCATTTTCAATGGAACGGGTGAATTAATCTCAACACTGAAATTTG 61591
QY 559 AsnGlnArgAlaValIleuGlnTrpArgSerSerAlaIysValAspAlaIleuTyryArg 576
Db 61592 AGGAGGAAAATTTTATGACGGCTTACTATGAAGAAATTGAATCTATGATCAG 61645

RESULT 12
ADL72353
ID ADL72353 standard; cDNA; 2103 BP.
XX

Db 568 TCTGGTACCAACGGGTGAGCCTTAAGCGGTATGCTTATACCGCAATTTAGCCTATCAG 627
Qy 207 LeuCyAlaAsnGlnMetLeuGlnThrPheProValPheGlyGluProPro 226
Db 628 TTGTATTCATGATCATCATCTCTGGCATTAATCT-----GATCAGAGAC 669
Qy 227 ValLeuValAspThrLeuProThrPheAsnIleThrPheGlyGlySerHisAsnIleGlyIle 246
Db 670 ATTCTTTGAGTTTCTGCGGTTATCCCATGCTTTTGAAGCGGTATGAGATGAGATCTTATGTC 729
Qy 247 ValLeuValAsnGlyGlyThrTyThrLeuAspAspGlyLysProThrAlaGlnGlyPhe 266
Db 730 ATGCATGTAGGGGACAAATAATCTTATCTGACGTGACACTAATTTGTCCTCAAGCATAG 789
Qy 267 AlGlnThrLeuArgAsnLeuSerGluIleSerProThrAlaTyThrValProLys 286
Db 790 GCG-----GATGCTCGCCGACAGTGAATGTCGATGCGCCGAGA 828
Qy 287 GlyThrGluGluLeuValGlyAlaLeu---GluArgAspSerThrLeuArgGluArgPhe 305
Db 829 TTTTACGAAAAATTTATAGTGGGTTTGAAGGTTCT-----AAGGCTCCT 879
Qy 306 PheAlaArgMetLysLeuPhePhePheAla---AlaAlaGlyLeuSerGlnGlyIle--- 323
Db 880 TGGCATGGAAGAACTTTTAAATGGCGCATTAATGTGGTGAAGCTCAGTTTAACT 939
Qy 324 -----TrpAsp---ArgLeuAspArgValAlaGluGln 333
Db 940 AAGTGCAGAGTAAAAAAGCTCCCTGGTCAACACATGATGATCAGCTGGATGATAA 999
Qy 334 -----HisCysGlyGluArgGlyLeuArgMetMetAla----- 343
Db 1000 TTGGTGTGAGCAAAATTACGAAATGTGTGGGGGAGAGTGGTGTAACTGCTGCT 1059
Qy 343 ----- 343
Db 1060 GGTGACAGTGTGACGAAAGATTAATCTGTTCTTCGCTGTTGGGATTAATGCAA 1119
Qy 344 ---GlyLeuGlyMetThrGluThrAlaPro---SerCysThrPheThrGlyPro 360
Db 1120 TATGGCTATGAGTGAAGAGACTTGTGCACAGGTTTCTGTTGGGAAGCAAGGT--- 1176
Qy 361 LeuSerMetAlaGlyTyThrIleGlyLeuProAlaProGlyCysGluValLysLeuValPro 380
Db 1177 ---TATGTACTTGTGTTATGTGAACCAATCGCTGCATGTGACGTCCTATCGTGA 1233
Qy 381 ValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyThrArg 400
Db 1234 GAAGAT-----GAGATACAAAGTCCGTGGCCCATTTGATGAAGGCTATTTCAAT 1284
Qy 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGlyTyThrCysSerGlyAspAla 420
Db 1285 AATCAGCAAGAAACCAACCAAGGCTTATCTCAAGACGTTGGTGGCTGACAGGATGCC 1344
Qy 421 IleLeuLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAla 440
Db 1345 GGTAAAGTTG-----GATGAAATGTAAATTTGTCATTACTGAACGCTGAAA 1392
Qy 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 1393 GATTGATGAAGAAACATGATGCAATATATTCCT-----CCTCAATGATGGAAGGC 1446
Qy 461 ValLeuGluGlyGlySerTyThrValLeuAspValValAlaAlaProAspArgGluCys 480
Db 1447 ATTTGGGACAAAGATCGTTTATGAAACATGTTGCTTATGCTGCTGCTGCTGCTGCTGCT 1506
Qy 481 LeuGlyLeuLeuValPhePro-----ArgLeuLeuAspCysArgAlaLeuSerGlyLeu 498
Db 1507 GTATCGGCTTTATCGTCTTGTGATGAAGCATGGAAGATGGAAGATGCAAGCTCAATTAAT 1566
Qy 499 GlyLysGluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaThrPheAla 518

Db 1567 CTCAAATATCATGACCGGCTGATTAATGCTGCTCATAGTCAGATATATATGTTGAG 1626
Qy 519 AspThrLeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetThrVal 538
Db 1627 CAACGCTTGAAAGAAATGAGAGAAAC---TTTGTCACTTCATGATGAGTAAAGCTTTC 1683
Qy 539 GlyLeuLeuAspThrProProSerIleAspLysGlyValThrAspLysGlySerIle 558
Db 1684 ATATTACTGCAAAACATTTTCAATGGAAGCGGTGAATTAATCCACACTGAATATG 1743
Qy 559 AsnGlnArgAlaValLeuGlnThrPargSerAlaLysValAspAlaLeuTyArg 576
Db 1744 AGAGAGAAATATTTATGTCAGCGTCTACTATGAAGAAATGAACTATGATCAG 1797

RESULT 11
ACF67367_48
Continuation (49 of 57) of ACF67367 from base 4800001 (Photobabdue luminescens nucleotic
WP Sequence split into 57 fragments Locus ACF67367 Accession ACF67367
WP Fragment Name Begin End
WP ACF67367_00 1 110000
WP ACF67367_01 100001 210000
WP ACF67367_02 200001 310000
WP ACF67367_03 300001 410000
WP ACF67367_04 400001 510000
WP ACF67367_05 500001 610000
WP ACF67367_06 600001 710000
WP ACF67367_07 700001 810000
WP ACF67367_08 800001 910000
WP ACF67367_09 900001 1010000
WP ACF67367_10 1000001 1110000
WP ACF67367_11 1100001 1210000
WP ACF67367_12 1200001 1310000
WP ACF67367_13 1300001 1410000
WP ACF67367_14 1400001 1510000
WP ACF67367_15 1500001 1610000
WP ACF67367_16 1600001 1710000
WP ACF67367_17 1700001 1810000
WP ACF67367_18 1800001 1910000
WP ACF67367_19 1900001 2010000
WP ACF67367_20 200001 2110000
WP ACF67367_21 2100001 2210000
WP ACF67367_22 2200001 2310000
WP ACF67367_23 2300001 2410000
WP ACF67367_24 2400001 2510000
WP ACF67367_25 2500001 2610000
WP ACF67367_26 2600001 2710000
WP ACF67367_27 2700001 2810000
WP ACF67367_28 2800001 2910000
WP ACF67367_29 2900001 3010000
WP ACF67367_30 3000001 3110000
WP ACF67367_31 3100001 3210000
WP ACF67367_32 3200001 3310000
WP ACF67367_33 3300001 3410000
WP ACF67367_34 3400001 3510000
WP ACF67367_35 3500001 3610000
WP ACF67367_36 3600001 3710000
WP ACF67367_37 3700001 3810000
WP ACF67367_38 3800001 3910000
WP ACF67367_39 3900001 4010000
WP ACF67367_40 4000001 4110000
WP ACF67367_41 4100001 4210000
WP ACF67367_42 4200001 4310000
WP ACF67367_43 4300001 4410000
WP ACF67367_44 4400001 4510000
WP ACF67367_45 4500001 4610000
WP ACF67367_46 4600001 4710000
WP ACF67367_47 4700001 4810000
WP ACF67367_48 4800001 4910000
WP ACF67367_49 4900001 5010000
WP ACF67367_50 5000001 5110000
WP ACF67367_51 5100001 5210000
WP ACF67367_52 5200001 5310000
WP ACF67367_53 5300001 5410000

OY	451	ValSerValGlyProLeuAaGfmrAaGAlaValLeuGluGlySerYrValLeuAsp	470
Db	10512	ATTTCGCTCAAGCGCATGACG-----AATTCGCTCAAGAAAGAC	10550
OY	471	-----ValValValAlaAlaProAspArgGluCysLeuGlyLeuVal	485
Db	10551	CCCCCTTCGAACACGCGGTGCTGGCGGAGATACCGTCCGGCTGACTGCTCGTT	10610
OY	486	PheProArgLeuLeuAspCysArgAlaLeu-----SerGlyLeu	498
Db	10611	AAGCCGTCGCTGCTCAGGTAGAGAGATTGGCCGAAGGTTGCATATCACTCAATGACG	10670
OY	499	GlyLySGluLaseSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAla	518
Db	10671	GGCCCGGAAMATGCTGCGCTTCGAGAGAGCTACGCGAGAA---ATTGTCGCCGGGTGGCC	10727
OY	519	AspTrpLeuLysArgLeuAsnArgGluValAlaThrGlyAsnAlaSerArgIleMetTrpVal	538
Db	10728	GAGATATCACCGAATAATGCTCCTCATCAAGACAGATATCGATCTCGCAAGTGTGGGGAC	10787
OY	539	GlyLeuLeuAsnSphrProProSerIleAspLySGlyGluValThrAspLySGlySerIle	558
Db	10788	GAGTTC-----ACCACAGACACGACGCTGCTCACTCAACCTCAAGGTA	10832
OY	559	AsnGlnArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaLeuTyr-ArgGly--	577
Db	10833	CGCCGACGTGAGGTGGAGAAAGATTCAACCGAGATGTTGAAGATATGACGGGGCTT	10892
OY	578	---GluAspGlnSerMetLeuArgAsp	585
Db	10893	GCGGCCCCGACGAAAGGCTGGTAAAGAG	10920
RESULT 10			
XX	ACF71666	standard; DNA; 1803 BP.	
XX	ACF71666;		
XX	20-NOV-2003	(first entry)	
Db	Photorhabdus luminescens nucleotide sequence #10133.		
XX	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;		
XX	detection; food; gene expression; plant; animal; microorganism; toxin;		
XX	antibiotic; biopesticide; virulence factor; disease model; plague;		
XX	whooping cough; gene; ds.		
XX	Photorhabdus luminescens.		
XX	WO200294867-A2.		
XX	28-NOV-2002.		
XX	07-FEB-2002; 2002WO-IB003040.		
XX	07-FEB-2001; 2001FR-00001659.		
XX	(INSP) INST PASTEUR.		
XX	(CNRS) CNRS CENT NAT RECH SCI.		
XX	Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunst F, Danchin A;		
XX	Buchrieser C;		
XX	WPI; 2003-148459/14.		
XX	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,		
XX	useful e.g. as therapeutic antimicrobials and agricultural pesticides.		
XX	Claim 2; SEQ ID NO 10133; 1205bp; French.		
XX	The invention relates to the isolation of genes and their encoded		
XX	proteins from Photorhabdus luminescens. The isolated sequences are		

	CC	sources of probes and primers for detecting the genome of P. luminescens
	CC	and related species; to study polymorphisms; for gene analysis and for
	CC	detection/amplification of the genes. Antibodies (Ab) raised against the
	CC	polypeptides encoded by the genes are used for detection/identification
	CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
	CC	carry a gene-containing vector are used to select compounds that
	CC	modulate, regulate, induce or inhibit expression of the genes in plants,
	CC	animals or microorganisms other than P. luminescens and are able to alter
	CC	response or sensitivity to toxins and antibiotics produced by P.
	CC	luminescens. Cells transformed to express the genes are useful for
	CC	recombinant production of the proteins, particularly toxins and
	CC	antibacterials useful as insecticides, bactericides and fungicides. The
	CC	genes, proteins, vectors containing the genes and Ab are also useful
	CC	therapeutically (to treat microbial infection by bacteria or fungi that
	CC	are sensitive to P. luminescens-encoded toxins or antibiotics) and as
	CC	biopesticides. Other uses of the genes and the proteins are as virulence
	CC	factors and for identifying targets of human diseases for which P.
	CC	luminescens is a model (particularly plague and whooping cough). This
	XX	sequence represents one of the isolated P. luminescens genes
SQ	Sequence 1803 BP, 502 A, 324 C, 423 G, 554 T, 0 U, 0 Other;	
	Alignment Scores:	
Pred. No.:	2,79e-16	Length: 1803
Score:	290.00	Matches: 146
Percent Similarity:	40.13%	Conservative: 110
Best Local Similarity:	22.88%	Mismatches: 264
Query Match:	9.41%	Indels: 118
DB:	10	Gaps: 26
US-09-750-986D-30 (1-589) x ACF71666 (1-1803)		
OY	11	ProGlyArgIleLeuSlnrLeuGluHisTrpAlaLysThrArgProGluGlnThrCys 30
Db	22	CCTTATACACTTGATGAACCGGTGGCAACAATAAACAACACCAGAGA----- 75
OY	31	ValAlaLarArgAlaLaIsnOlyLutTrpArgHrgLieserTryAlaGluMet---Phe 49
Db	76	ATTGCCTTTTCG-----CAATGGTCTCGCGCTGAACAACCTGGAACTTAAGTGG 123
OY	50	HlaAsnVal-----ArgAlaIleAlaInsIserLeuLeuProTryGlyLeuSer 65
Db	124	CATGATATTGGCAAAGAAGCSAGCCTATTGACGGCGCTTTGGAGATTGGGTGAT 183
OY	66	AlaGlnArgProLeuLeuIleValSerGlyAsn-----AspLeu 78
Db	184	GTGCAGAAAATAATGGCTCTTTGCACATGAACAGCATGGCGTGCATTGGCAGATATC 243
OY	79	GlnHisLeuSlnrLeuAlaIleAspLeuLysMetTryAlaGlyIleProTryCysProValSer 98
Db	244	GCTGATTTACAACTTTCG-----GGCGTCACTGTC 273
OY	99	ProAlaTrySerIleLeuSerGlnIleAspLeuAlaLysLeuArgHisIleValGlyLeuLeu 118
Db	274	CTCTGTGATGAACCAACGATGTGACGAACAGCGTGGTTC-----ATACTTAATGATGCT 327
OY	119	GlnProGlyLeuValPheAlaIleAspAlaIleAspProGlnArgAlaIleGlu--Thr 137
Db	328	GGTGTGCGATTTGTTGTTGTGTGGCGACGCGCAGATGATGTTGCTATTAACGTTAC 387
OY	138	IleLeuProAspArgValProAlaIlePheThrArgGlyGluLeuAlaGlyArgGlyThr 157
Db	388	GAAATGTTGCCCTCAGTTAAGTCACATTTGTCTTAGATGAATCAGTCGATTTGCCGGAT 447
OY	158	ValSerPheAspSerLeuLeu-----GluGlnProGlyGlyIle 170
Db	448	TGTTCTATTAGCTCAGCATTTATCTGATTTCAATGSCATGAATGAATCACTGTCACTC 507
OY	171	GluAlaAspAspAlaPheHeAlaIleThrGlyProAspThrIleAlaLysPheLeuPheThr 190
Db	508	GAGTGGAAACCGATATGCTAACCAACAATCGGATGATCTTTTAACATTAATTATACG 567
OY	191	SerGlySerThrLysLeuProLysAlaValProThrThrglnArg-----Met 206

QY 113 HisIleValGlyLeuLeuGlnProGlyLeuValPheAlaIleAspAlaIleProPheGln 132
 Db 9396 CACATCGTCACTGATGCTGGGGTGCCTGTCATCAACGGCTGGTCCCAAGAGACTGCAC 9455
 QY 133 ArgAlaIleGlnThrIleLeuProAspAspValProAlaIlePheThr----- 148
 Db 9456 CGTATTCGAGGCC-----CGCGATCAAGTCCGGCCCTTGAGCGGTGATCTGTATA 9509
 QY 149 -----ArgGlyGlyLeuAlaGlyValArgThrValSerPheAspSerLeu 163
 Db 9510 AATCCGACATCAAGTGGTGGTGCATGACGCGTTTGACGCGTCTGCTCCGGAACAG--- 9566
 QY 164 LeuGlnGlnProGlyGlyIleGly-----AlaAspAsnAlaPheAla 178
 Db 9567 GTGGCTGACAGCTGGGGTTCGAGAGATACAGACGCTCGTCCAGAGCGCATGGGCGAG 9626
 QY 179 ThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuProLys 198
 Db 9627 TCATGCCCGACGATGTGGCCGCTGTGATTACCTTGGAACACCGCGGTGACCCCAAG 9686
 QY 199 AlaValProThrThrGlnArgMetLeuCyAlaAsnGlnIleMetLeuGlnThrPhe 218
 Db 9687 GGGGTCTGATCAACCCATCCGGCCGCTGGCTAG-----CTGCAGGCACTT 9734
 QY 219 ProValPheGlyGlnGluProProVal-----LeuValAspTrpLeuProTrpAsnHis 236
 Db 9735 GAGCGCTTCTTCGACGTCACCCGAGTCGATCACTCGTAGACTTCTGCGCGCTCTCCAC 9794
 QY 237 ThrPheGlyLysSerHisAsnIleGlyIleValLeuTyraAsnGlyThrTyrrTyLeu 256
 Db 9795 GCCCTGAGTGGGAGATGGTATGGCGGTATTCGATTCGATTCCTCAATACCTTCGTG 9854
 QY 257 AspAspGlyLysProThrAlaGlnGlyPheAlaGlnThrLeuArgAsnLeuSerGlnIle 276
 Db 9855 CCCAACCCGAAAGCATCTG-----GCATGTTGGCTGAGGTC 9883
 QY 277 SerProThrAlaTyrrLeuThrValProLysGlyTrpGlnGluLeuValGlyAlaLeuGlu 296
 Db 9894 CGCCCAACCTTTCGTCGTCAGTCCGAACTGTACGAAACAAGTCATG----- 9941
 QY 297 ArgAspSerThrLeuArgGluArgPhePhe-----AlaArgMetLysLeuPhePhe 313
 Db 9942 -----AGCGTGGCCGTGAGAGGTTTCTGATTCCTCCGCAAGCTCAAAATCTTTGAG 9995
 QY 314 PheAla----- 315
 Db 9996 TGGTCATATACGATGGCCCGAGTGGTGGCAAGCCGACGAGGAGGACGGCCGCCAC 10055
 QY 316 ---AlaAlaGlyLeuSerGlnGlyIleTrpAspArgLeu----- 327
 Db 10056 GTATCTTTGGAGAGGCGTCATGGGGTTGCCGATGTCGTCCGTCCAGAGGCATCCGAC 10115
 QY 328 ----- 330
 Db 10116 GCCATCGGTGAGCCAGACTGCTTGCAGCCGCGGAGCTCCATCTCCGCAAGAGACT 10175
 QY 331 AlaGlnGlnHisCysGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrThr 350
 Db 10176 GAAGAGTTCTTCCCGCTCGGCTGTTGGTGGCCAGGAGATACGGGCTGACGAGAGGCC 10235
 QY 351 AlaProSerCysThrPheThrGlyProLeuSerMetAlaGlyTyrrIleGlyLeuPro 370
 Db 10236 TCCCGGTGGTAGCTTCACTCTCCAGGGGATACAAAGTTCCGCACTGGGAGAGCCG 10295
 QY 371 AlaProGlyCysGluValLysLeuValProValAspGlyLysLeuGlnGlyArgPheHis 390
 Db 10296 CTAGTGGGTGATGATGACACAGACTGAG---GACGGGAGATTT-----CTCTAACGTT 10346
 QY 391 GlyProHisValMetSerGlyTyrrTrpArgAlaProGlnGlnAsnAlaGlnAlaPheAsp 410
 Db 10347 GGTCCCAAGCTCATGAAGGATTTGGAAGCCCTGAGGCCACTCGCCGACGCAATC-- 10403

QY 411 GlnGlnGlyTyrrTyrrCysSerGlyAspAlaIleLysLeuAlaAspProAlaAspProGln 430
 Db 10404 GAAGATGCTGTTGTTCACACTGGGGATATTGGCACATC-----GACGAAGAT 10451
 QY 431 LysGlyLeuMetPheAspGlyArgIleAlaGlnAspPheLysSerSerGlyValPhe 450
 Db 10452 GGTTTTCTGTCATTACCGATCCGCTCAAGACATCATCTCATCCCTTAACGGCAAGAT 10511
 QY 451 ValSerValGlyProLeuArgThrArgAlaValLeuGlnGlySerTyrrValLeuAsp 470
 Db 10512 ATTTCGCTAGCCGATCGAG-----AATTGCTCATGAAGAGAC 10550
 QY 471 -----ValValAlaAlaProAspArgLysCysLeuGlyLeuVal 485
 Db 10551 CCCCTTCAGACACGGCGGTGCTGGCGGATTAACCGTCCGTGCTGACTGCTGCTT 10610
 QY 486 PheProArgLeuLeuAspCysArgAlaLeu-----SerGlyLeu 498
 Db 10611 AAGCGTGGCTGCTCAGTGAAGAGTTGGCCGAGGTTGCATATCACTCATAGACG 10670
 QY 499 GlyLeuGlnLaseerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAla 518
 Db 10671 GCGCCGAAATGCTGCGTTCGAGAGACTAGCCGAGAA---ATTGTCGCGGGTGGCC 10727
 QY 519 AspTrpLeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpVal 538
 Db 10728 GAGATCACCGAAGATTGCTCATCAAGACAGATTGCTGATCTGCAGAGTGTGGGAC 10787
 QY 539 GlyLeuLeuAspThrProProSerIleAspLysGlyGlyValThrAspLysGlySerIle 558
 Db 10788 GAGTTC-----ACCACAGACACGAGTGTCTCACTCCAACCTTCAGGTA 10832
 QY 559 AsnGlnArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaLeuTyrrArgGly-- 577
 Db 10833 CGCCGACGTAGGTGGAAGAGATTCACCGAATCTTGAAGAGTGAACGGCGGCTT 10892
 QY 578 ---GluAspGlnSerMetLeuArgAsp 585
 Db 10893 GCGGCCGACGAAGGCGTGTAAAGAG 10920
 RESULT 9
 ACF64451
 ID ACF64451 standard; DNA; 37286 BP.
 XX ACF64451;
 AC
 XX 17-OCT-2003 (first entry)
 DT
 XX
 DE Propionibacterium acnes DNA contig sequence #17.
 KW Acne vulgaris; antiseportheic; dermatological; antibacterial;
 KW immunostimulant; immune response; vaccine; ds.
 XX
 OS Propionibacterium acnes.
 XX
 XX
 PN W0200303515-A1.
 PD 24-Apr-2003.
 XX
 PF 11-OCT-2002; 2002W0-US032727.
 XX
 PR 15-OCT-2001; 2001US-00978825.
 PA (CORI-) CORIXA CORP.
 XX
 PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL,
 PI Zhang Y, Wang S, Jen S, Lodes WD, Benson DR, Jones R, Carter D;
 PI Barth B, Valiieve-Douglas J;
 XX
 DR WPI; 2003-381769/36.
 XX
 PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
 polypeptide, useful for diagnosing, preventing or treating acne vulgaris,

```
QY 385 LeuGIUGLYArpHe-----HisGIYProHisValMetSerGIYTrpArgAlaPro 402
Db 33900 ACCGTTGGAGGGCTTACAGTTCCTAGTCCAGACGATGTCGGCGGCTATCTAAATCCAGCG 33959
QY 403 GIUGIAsnAlaGlnAlaPheAspGIUGLYTrpYrCysSerGIYAAspAlaIleLys 422
Db 33960 GAAGCCAGACGCGAGCGGCTTCAGAGAGAGCGTTGGTACGTTAGTGTAT---GTTGCG 34016
QY 423 LeuAlaAspProAlaAspProGlnLysGIYLeuMetPheAspGIYArgIleAlaLysAsp 442
Db 34017 GTCTGAGACAGTGGCGGATGATCCG-----ATCGTGGGTCCGAGTCGGTTGAC 34067
QY 443 PheLysLeuSerSerGIYValPheValSerValGIYProLeuArgTrpArgAlaValLeu 462
Db 34068 CTAATCAATATGGCGGATACCGGATCCGCGCAGAGAAATC-----GAGATGGCGCTG 34121
QY 463 GIUGIYGIYSerYrValLeuAspVal---ValValAlaAlaProAspArgGIYCyLysLeu 481
Db 34122 CTAGGACATCCGATGTCGGGAGCTGCGAGCTTGTCCGGCTGCTGAC---GAGGATTTG 34178
QY 482 GIYLeuLeuValPheProArgLeuLeuAspCyArgAlaLeuSerGIYLeuGIYLYSGIY 501
Db 34179 GGTCAACGCGATCGTTCGATTCGTC-----GTAGGCGGCAGAG 34214
QY 502 AlAspAspAlaGIYValLeuAlaSer-----Glu 511
Db 34215 GCATTGATGCTCCGATGATGATTAACTATGTTGCCACAGACCTTTCGATACATAGACGG 34274
QY 512 ProValArgAlaTrpPheAlaAspTrpLeuLysArgLysAsnArgIleAlaThrGlyAsn 531
Db 34275 CCGCGTAACTCCGCTTCGTGAGACGCGCTCCGCGC---AAGCG-GATGGGAAAGTCTT 34330
QY 532 AlAspArgIleMetTrpValGIYLeuLeuAspTrpProSerIleAspLysGIY 551
Db 34331 CAAAGACAGTGTCTCTTCGAGGCGTACTGCGCTACGACGAGTATGATATGCGCGC 34390
QY 552 ValThrAspLysGIYSerIleAsnGlnArgAlaValLeuGlnTrp-----566
Db 34391 ACCGCATCGTTCGACCGGACCGGCGGCGCGGATGTGTTGGTTCGCGCGCCGCTACTGT 34450
QY 567 ArgSerAlaLysValAlaPheAlaLeuTrp-----ArgGIYGIYAspGlnSer 581
Db 34451 AGAGCTTCAAGACGACGACGACATGTGACCGTACTGTGTGTCAGAGGCTCCGGTAGAGAT 34510
QY 582 MetLeuArgAsp 585
Db 34511 TTTTCAGCGGAC 34522

RESULT 8
AAS59522 standard; DNA; 37286 BP.
XX AAS59522:
AC
XX
XX AAS59522:
DT 13-FEB-2002 (first entry)
XX
XX Propionibacterium acne immunogenic protein encoding DNA #17.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteoporotic; neuroprotectant; ds.
XX
XX Propionibacterium acne.
XX
XX PN MO200181581-A2.
XX
XX PD 01-NOV-2001.
XX
XX PF 20-APR-2001; 2001MO-USO12865.
XX
XX PR 21-APR-2000; 2000US-0199047P.
```

```
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
XX (CORI-). CORIXA CORP.
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhattacharya A;
XX L'Amourneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
XX
XX Propionibacterium acne polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris.
XX
XX Claim 1; SEQ ID NO 17; 1069pp; English.
XX
XX Sequences AAS59506-AAS59804 represent DNA molecules encoding
XX Propionibacterium acne immunogenic polypeptides. The proteins and their
XX associated DNA sequences are used in the treatment, prevention and
XX diagnosis of medical conditions caused by P. acnes. The disorders include
XX SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
XX osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in
XX infections of bone, joints and the central nervous system, however it is
XX particularly involved in the inflammatory lesions associated with acne
XX vulgaris. A method for detecting the presence or absence of P. acnes in a
XX patient comprises contacting a sample with a binding agent that binds to
XX the proteins of the invention and determining the amount of bound protein
XX in the sample. The polypeptides may be used as antigens in the production
XX of antibodies specific for P. acnes proteins. These antibodies can be
XX used to downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA). This sequence encodes the
XX polypeptides shown in AA044158-AA044568 and AA067488-AA067490. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences

Alignment Scores:
Pred. No. 1.52e-15 Length: 37286
Score: 301.50 Matches: 154
Percent Similarity: 39.10% Conservative: 108
Best Local Similarity: 22.99% Mismatches: 264
Query Match: 9.78% Indels: 144
DB: 4 Gaps: 29

US-09-750-986D-30 (1-589) x AAS59522 (1-37286)
QY 9 ProPheProGIYArg-----IleLeuGlnValLeuGlnIleTrpAlaLysTrp----- 24
Db 9066 CCGTACCGGGGAGAACTTCAATGAGACCACTGCGCCACATGTTCCGCGCCAGCGTT 9125
QY 25 -----ArgProGlnGlnTrpCysValAlaAlaArgAla---AlaAsnGIY 38
Db 9126 GCCAACCAAGGCTTCGCTCC-----GCCACTCGGCTGTCAGGCGCGT 9170
QY 39 GIUTrPArgArgIleSerYrAlaGlnMetPheHisAsnValArgAlaIleAlaGlnSer 58
Db 9171 CAGTGATTTATCCGACATGCGAGACGTGGGCGTGGCTACCGGCGCTCGCGCGGCC 9230
QY 59 LeuLeuProTrpGIYLeuSerAlaGIYValGrgProLeu-----LeuIle 72
Db 9231 TTGTGCACTCCCGGGGTGCTACCGAGAGATGGCTCAGCGCGGAGATGATTTCTCTG 9290
QY 73 ValSerGIYAsnAspLeuGlnIleLeuGlnIleValPheGIYAlaMetYrAlaGIYIle 92
Db 9291 TTGCGCGGATTTGCGCGAGATGATCGAGCGATGCTCGCGGCGATGAGATGGCGCGT 9350
QY 93 ProTrpCyAspProValSerProAlaTrpSerLeuLeuSerGlnAspLeuAlaLysLeuArg 112
Db 9351 -----ATCCCGGCTG---CCGATTTACCGAGCGTCAACCCCGTAT-----CAGATTGTT 9395
```

Db 18317 CACGCCATCGTTGGACCGGACGGGGCCGGATGTGTGGTGGCGCCGCCCACTG 18376
|||||
Qy 567 -ArgSerAlaIysValAspAlaLeuTyr-----ArgIyGIuaepGlnse 581
|||||
Db 18377 TAGAGCTTCAAGACCACCGCAATGTGGACCGTAGTGTGTGTCAGAGGGGTCCGGGTAGAG 18436
Qy 581 rMetIeuArgAsp 585
|||||
Db 18437 TTTTTCAGCGGAC 18449
RESULT 7
ADB74389
ID ADB74389 standard; DNA; 35961 BP.
XX ADB74389;
XX
XX 04-DEC-2003 (first entry)
XX
XX Mycobacterium leprae DNA #23.
XX
XX Non-naturally occurring peptide; anion pump protein; tuberculosis;
KM hypersensitivity reaction; tuberculostatic; gene; ds.
XX
XX Mycobacterium leprae.
XX
XX US6583266-B1.
XX
XX 24-JUN-2003.
XX
XX 16-SEP-1994; 94US-00311731.
XX
XX 19-AUG-1993; 93US-00109181.
XX 22-OCT-1993; 93US-00142558.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Smith DR, Mao J;
XX
XX WPI; 2003-656441/62.
XX
XX New Mycobacterium tuberculosis anion pump peptide useful for as
PT tuberculosis vaccine and diagnosis of tuberculosis infection.
XX
XX
XX Disclosure; SEQ ID NO 138; 26pp; English.
XX
XX The invention relates to a non-naturally occurring peptide of
CC Mycobacterium tuberculosis comprising an amino acid sequence
CC corresponding to an anion pump protein. The invention also relates to a
CC non-naturally occurring nucleic acid corresponding to a DNA sequence of
CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
CC leprae or for screening for new tuberculosis drugs. Purified proteins
CC derived from the sequences of the invention may elicit a specific immune
CC response. The peptide may also be used to detect hypersensitivity
CC reactions of individuals exposed to Mycobacterium tuberculosis or
CC Mycobacterium leprae. The proteins and peptides may be affixed to solid
CC supports to detect antibodies typical of hypersensitivity reactions, from
CC a patient's sera. This sequence represents Mycobacterium leprae DNA of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format
CC directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX
XX Sequence 35961 BP; 6748 A; 9918 C; 11140 G; 8155 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1.18e-15 Length: 35961
Score: 302.50 Matches: 150
Percent Similarity: 37.50% Conservative: 69
Best Local Similarity: 25.68% Mismatches: 234
Query Match: 9.82% Indels: 132
Gaps: 27

US-09-750-986D-30 (1-589) x ADB74389 (1-35961)
Qy 44 SerTyrAlaIuMetPheHisAsnValArgAlaIleAlaIleInsLeuLeuProTyrGly 63
|||||
Db 33039 TCGGTGGCTAGACGCGTGTGCGAGGTGCACGCGCTGTGTACGGGTGCGCACGCCG----- 33092
Qy 64 LeuSerAlaIuArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeu 83
|||||
Db 33093 ---ACGCGCATCGACTGTGTAGCGATTACT----- 33119
Qy 84 AlaPheGlyAlaMetTyrAlaIleProTyrCysProValSerProAlaTyrSerLeu 103
|||||
Db 33120 ---GGTTCCTGATTGCCGGGTGTCGGGTGTCCGGGTGCCCGGATATCGGTGTG 33173
Qy 104 LeuSerGlnAspLeuAlaIysLeuArgHisIleVal-----GlyLeu 117
|||||
Db 33174 GTCGAAAGC-----CGGCATATGCTACACCGATTCTGGGGCGGACGCTGG 33218
Qy 118 LeuGlnProGlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThr 137
|||||
Db 33219 CTGGGTCCGGGGCC-----AGTCCGACTCCGCG----- 33248
Qy 138 IleLeuProAspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgThr 157
|||||
Db 33249 ---CCGACGGGTGGCCGCATATC-----CCG 33272
Qy 158 ValSerPheAspSerLeuLeuGlnProGlyGlyIleGluAlaAspAsnAlaPheAla 177
|||||
Db 33273 GTTCAATAGATGCT-----CGCTCTGGAAACCGTATCCT 33308
Qy 178 AlaThrGlyProAspThrIleAlaIysPheLeuPheThrSerGlySerThrIysLeuPro 197
|||||
Db 33309 GAGCGGTCACTGATGACCGCGATGTGATGATCAACCTCGGGTACTACTGGGCCCT 33368
Qy 198 LysAlaValProThrThrGlnArgMetLeuCysAlaIasnGlnMetLeuGlnThr 217
|||||
Db 33369 AAGGGGTGATGTTGACCGCGGACGATGCGGTGATTGGAACGCGTTGGCCAGGCC 33428
Qy 218 PheProValPheGlyGluGluProProValIleuValAspTyrLeuProTyrAsnHisThr 237
|||||
Db 33429 TGGCAGTGAACGCCCATGAC-----GTTTGTATCATGATTTGCCGCTTTCAAGCTG 33482
Qy 238 PheGly-----GlySerHisAsnIleGlyIleValLeuTyrAsnGly 251
|||||
Db 33483 CACGGCTTGACTCGGTTGCTTGATGCTTCGAT----- 33521
Qy 252 GlyThrTyrTyrLeuAspAspGlyLysProThrAlaIleGlnIysPheAlaIuThrLeuArg 271
|||||
Db 33522 GGAATGCTTGTGTACACACCGAAMAAACCATCCGACCGCTATGCTCAGGCT----- 33575
Qy 272 AsnLeuSerGluLeuSerProThrAlaTyrLeuThrValProLysGlyTyrGluGluLeu 291
|||||
Db 33576 ---TGCTTTAAGCCGGGGATCTCTGTATTTGGCGCTTCCACGATGATGCGGCTG 33632
Qy 292 ValGlyAlaLeuGluArgAspSerThrLeuArgGluArgPheAlaArgMetLysLeu 311
|||||
Db 33633 GTGGCA-----GACGAGGACAGT---GCCGGCGGTAAAGACCGCGGCTA 33677
Qy 312 PhePhePheAlaAlaAlaGlyLysSerGlnGlyIleThrAspArgLeuAspArgValAla 331
|||||
Db 33678 CTAGTGTCTGGAGTGTGCTCATGCGCGTTCGCGTATTTGATGCGGTAGCGCACTTACA 33737
Qy 332 GluGlnHisCysGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAla 351
|||||
Db 33738 GGTTCATCGCCCATCGAACG-----TATGGACAGACGGAATTCATCTG 33779
Qy 352 ProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAla 371
|||||
Db 33780 ATCAACCTGTGCAACGTTGGGGAACGAGACGCGCGTGGCGGTGGGATTTGCCGCTG 33839
Qy 372 ProGlyCysGluValLysLeu-----ValProValAspGlyLys 384
|||||
Db 33840 GCGGGCGGTGCAGACACAGCTGTGTCATGATCGGAGGTCCAGTGCCTTACGACGGGAA 33899


```

Db 54975 TCGGCGCCAGGCGCTGATCGCGCTCGCATCGCAAGGCGACGGCGGTGACCATCTT 55034
Qy 74 SerGlyAsnAspLeuGluHisLeuGluLeuAlaPheGlyAlaMetYrAlaGlyLeuPro 93
Db 55035 TCCTCCACCCCGCTGAGATGGGGGATCTTACATTTCCACTGGCGCGGCGGCGCGGTA 55094
Qy 94 TyrCysProVal-----SerProAlaTyrSerLeuLeuSerGlnAsp 107
Db 55095 AGCGTCGCGATCTATGACACCGACTCCGCCCCGACGCGCCATCATGAAGATTCG 55154
Qy 108 LeuAlaIleuAspGluHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAsp 127
Db 55155 GTGGTCAAG-----CTGGCTTTGGAAAGCGCTAGGCGTGACC 55181
Qy 128 AlaAlaProPheGlnArg-----AlaIleGluThrIle 138
Db 55182 CGGAGGCGCTTCGACCGCGCTCGACTCGTCAAGACCATCGCTCGCTTGAACCAATT 55241
Qy 139 LeuProAspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArg---Thr 157
Db 55242 CTCATGATCGAGGCGACGCGCTG-----GGCGCTTTGGAAAGCGCTAGGCGTGACC 55292
Qy 158 ValSerPheAspSerLeuLeuGlnGlnProGlyGlyIleGluAlaAspAlaPheAla 177
Db 55293 GTTTCGACGAGAACTGACAGACGCGGTGGCCACCGTGGCGACGAGAC----- 55343
Qy 178 AlaThrGlyProAspThrIleAlaIlePheLeuPheThrSerGlySerThrIleLeuPro 197
Db 55344 -----CTGGCCACCATGCTCTACACTTCGAGGCTCCACCGGCAACCCG 55385
Qy 198 LysAlaValProThrThrGlnArg-----MetLeuCyAlaAsnGln 212
Db 55386 AAGGGTCCGAACTGACCCACAGAACTTGTCATCATCATCTCGCGCTGACAGCG 55445
Qy 213 MetLeuLeuGlnThrPheProValPheGlyGluGluProProValLeuValAspTrpLeu 232
Db 55446 TTG-----CATGAAGTGCTGCTTACGACCATCGCGCTCGCTCTCTCTG 55493
Qy 233 ProTrpAsnHisThrPheGlyGlySerHisAsnIleGlyIleValLeuTyrAsnGly 252
Db 55494 CCGCTGCGCCACGCTTCGCGCGCATTCATCGACGACCTCCATCGCTCGCATGACGCG 55553
Qy 253 ThrTyr---TyrLeuAspAspGlyLysProThrAlaGlnGlyPheAlaGluThrLeuArg 271
Db 55554 GTGGTCGATTATCTGCGCGACACCAA-----ACCGTCTGCGCG 55592
Qy 272 AsnLeuSerGluIleSerProThrAlaTyrLeuThrValProLysGlyTrpGluLeu 291
Db 55593 GACCTGGTTCGTTGAACCGACTACCTGCTGGCGCTGCGCGCTGCGAGAAAGTA 55652
Qy 292 ValGlyAlaLeuGluArgAspSerThrLeuArgGluArgPhePheAlaArgMetLysLeu 311
Db 55653 TACACCGCCCGCTCGCACAGGCGCGT-----GCCGCTGGAAAGGC 55694
Qy 312 PhePhePheAlaAlaAlaGlyLeuSerGlnGlyIleTrpAspArgLeuAspValAla 331
Db 55695 CGCCTGTTTGTCAAGCGCGCGAGCGCGCGCTGGAAGCGCAAGACGAGCGCGC 55754
Qy 332 GluGlnHis----- 334
Db 55755 GACGACGACACGTTTCCGAATCGCAGCGCGCAAGACGAAACGCTGCTACCGC 55814
Qy 335 -----CysGlyGluArg----- 338
Db 55815 ACGGTGCGCGGTGCTGCGCCGCAAGATCAAGTACGTGCGCGCGCGCGCACCATG 55874
Qy 339 -----IleArgMetAlaGlyLeuGly 346
Db 55875 TCGCTGATCTGGGCGATTCTACACGCGATCGGATCGCGATGATTCAGGGTTATGCG 55934
Qy 347 MetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyr 366
Db 55935 ATGACGGAACCGCGCGCTTGCAGCGCAAGCGCGTGAACGCAATGTGATCGGCGACC 55994

```

```

Qy 367 IleGlyLeuProAlaProGlyCysGluValLysLeuValProValAspGlyLysLeuGlu 386
Db 55995 GTAGGCGACCGCGCGCGCTCGCTCGATTCG---ATTCCGACGAAAGCGAACTGACG 56051
Qy 387 GlyArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAla 406
Db 56052 GTCAG-----GGCCGAAACGTTTCCGCGGTACCAACATCTGCGGGAAGACGCGC 56105
Qy 407 GluAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleLysLeuAlaAspPro 426
Db 56106 GAGCGTTTCACCGCCGACGCTGCTGAGACCGGCGACTTGGCCGAGATC----- 56156
Qy 427 AlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLysSer 446
Db 56157 ---GATGATGAGGCGCAATCGATCATCACCGCGCGATCAAGACATTAATTATTCGCC 56213
Qy 447 SerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGluGlySer 466
Db 56214 GCGCGCAAGAACGTCCTCCCATTCGCTGAGAAAG-----ATCGCAAGTGGCCG 56267
Qy 467 TyrValLeuAspValValAlaAlaProAspArgGluCysLeuGlyLeuValPhe 486
Db 56268 ATCGTGAGCATGCGGTTGGTGGGCGACGACGCGCGCTTCATCGCGCGTTGATG--- 56324
Qy 487 ProArgLeuLeuAspCysArgAlaLeuSer-----GlyLeuGlyLys 500
Db 56325 -----ACGCTCGATCCGAGCGTGGTGGCATTTGGCTGCCGCGACGCGCTGTCCAGC 56378
Qy 501 GluAlaSerAspAlaGluValLeuAlaSerGlu-Pro-----ValArgAlaTrpPheAl 518
Db 56379 GAG---ACGCGGTGACGCGCTCGCCACGAAACGCGCGTGGCGGAGATTCAGCACG 56435
Qy 518 AspTrpLeuLysArgTrpLeuAsnArgGluAlaThrGlyAsn-----AlaSerArgIle 536
Db 56436 TACGTGACAGGCGAATGACGACGTCTCGCGCGCGAATCGTACGCAAGTTGCGCGTG 56495
Qy 536 Trp 537
Db 56496 CTGG 56499

RESULT 6
ADB74275
ID ADB74275 standard; DNA; 38494 BP.
XX
AC ADB74275;
XX
DT 04-DEC-2003 (first entry)
XX
DE Mycobacterium leprae DNA #2.
XX
KW Non-naturally occurring peptide; anion pump protein; tuberculosis;
  hyperbemicity reaction; tuberculosis; gene; dr.
XX
OS Mycobacterium leprae.
XX
PN US6583266-B1.
XX
PD 24-JUN-2003.
XX
PF 16-SEP-1994; 94US-00311731.
XX
PR 19-AUG-1993; 93US-00109181.
XX
PR 22-OCT-1993; 93US-00142556.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Smith DR, Mao J;
XX
DR WP1; 2003-656441/62.
DR P-PSDB; ADB74303, ADB74304, ADB74305, ADB74306, ADB74307, ADB74308,
DR ADB74309, ADB74310, ADB74311, ADB74312, ADB74313, ADB74314, ADB74315,
DR ADB74316, ADB74317, ADB74318, ADB74319, ADB74320, ADB74321, ADB74322,

```



```

Db      702 GGACTGCTCTTGGCCGATGACGACCTTTCAG----- 737
Qy      141 AspaerylProalailepethrArgGlyGluLeuValaGlyArgThrValSerPhe 160
Db      738 -----GGGAGAGG----- 746
Qy      161 AspSerLeuLeuGluGlnProGlyGlyLeuValaAspaenalaPhealaIaThrGly 180
Db      746 ----- 746
Qy      181 ProasphrIlealaLysPheLeuPheThrSerGlySerThrGlyLeuProLysAlaVal 200
Db      747 -----CGGTTTGGCTATGCGGCATGATCAAAAACCTGTGTAATTCATT 791
Qy      201 ProthrThrGlnArgMetLeuCyAlaAsnGlnGlnMetLeuLeuGlnThrPheProVal 220
Db      792 AAGCATTCCTCCGACATGAAAGCCATCAAAACGCGCATGATGAA-CCTGAATCCGACGCG 850
Qy      221 PheGlyGluGluProProValLeuValAspTrpLeuProTrpAsnIsthrPhe----- 238
Db      851 GCATCAGCACCTTGCCCTGGCTATTAATTTGCCATGAGCAGCACCGTGAAACG 910
Qy      239 ---GlyGlySerHisasnIleGlyIleValLeuTyraenGlyGlyThrTyTrpLeuAsp 257
Db      911 GATGAAAGCAGAACCCAGTTCATGACCTGTTCCGTTCCGTAACTGTA----- 961
Qy      258 AspGlyLysProThrIleGlnGlyPheAlaGluThrLeuArgLeuSerGluIleSer 277
Db      962 -----ATGCAAGTAGCGTATGCGCTACCGCACTGGTCCAGAACCTTGACCGAACCGCAG 1015
Qy      278 Pro-----ThrAlaTyLeuThrVal 284
Db      1016 CGGTGTAAACGGCGCAGTGGCGTTTTCATGCGCTGTATGACGTTTTCATGACAGTC 1075
Qy      285 ProLysGlyTrpGluGluLeuValGlyAlaLeuGluArgAspSerThrLeu----- 301
Db      1076 TATGCTCGGCGCATCCAAAGCAGCAGCGCTTACGCGGTGATGATGATGATGATGATG 1135
Qy      302 ---ArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaIleGlyLeuSer 320
Db      1136 GAGCAGCAGACGATGTTACGACGACGACGACGATGTTACGACGACGAGGAGTCCCTTAAAA 1195
Qy      321 GlnGly-----IleTrpAspArgLeuAspArgValAla----- 331
Db      1196 CAAAGTAGGAGGCTCAAGATGAGGATTCATGCGACATGAGGCTGGCCCTGACCAAG 1255
Qy      332 GlnGlnHisGlyGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAla 351
Db      1256 TCAATTCATGCGG-----CTGCTCTGATCTTTTCGTCGTGAGTTCGAGACCTAG 1309
Qy      352 Pro-----SerCyThrPheThrThrGlyProLeuSerMetAlaGly 365
Db      1310 CCACTTACTCCCAACATCAGCCGACTCCGATTAACCTCGGGAAGCTCTCGTAAGA 1369
Qy      366 TyrIleGlyLeuProAlaProGlyCyGlyValLysLeuValProValAspGlyLysLeu 385
Db      1370 CATTCATCGGCTTGCGCTTCGACCAAGAGCGTTGTTGGCGCTC---TCCGCGCTT 1426
Qy      386 GlnGlyArgPheHisGlyProHisValMetSerGlyTyTrpArgAlaProGluGlnAsn 405
Db      1427 ACG----- 1429
Qy      406 AlaGlnAlaPheAspGluGluGlyTyTrpCySerGlyAspAlaIleLysLeuAlaAsp 425
Db      1430 -----TTCGCCACAGT----- 1441
Qy      426 ProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLysLeu 445
Db      1442 -----TTG 1444
Qy      446 SerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGluGlyGly 465

```

```

Db      1445 AGCAGCCGGTAGATGATCTATATCTATGATCTCGAGTCTCCGGCGAGACCGGAGGC 1504
Qy      466 SerTyraValLeuAspValValAlaAlaProAspArgGluCyLeuGlyLeuLeuVal 485
Db      1505 AGG-----GCATTGCCACCGGCTTCATCAATCTCTCAAGCATG 1543
Qy      486 PheProArgLeuLeu-----AspCyArgAlaLeuSerGlyLeuGlyLys 500
Db      1544 AGCCCAAGCGCTTGCTTATGTGATCTACGTCAGACGAGATTACG----- 1591
Qy      501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
Db      1592 -----GTACGATCCGCGAGTGG 1609
Qy      521 LeuTyraArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpVal----- 538
Db      1610 CTC-----TCTATACAACTTGGCGATTACGGG 1656
Qy      539 -----GlyLeuLeuAspThr-----ProProSerIleAspLysGlyGlu 551
Db      1637 AAGAAGTGATGCACTTGATATGACCCCAAGTACCGCACCTAAC----- 1681
Qy      552 ValThrAspLysGlySerIleAsnGlnArgAla-----ValLeuGlnTrpArgSer 566
Db      1682 -----AATTGTTCAAGCCGAGATCGGCTTCCCTGTTTGCATGCGGCTCG 1729
Qy      569 AlaLysValAspAlaLeuTyraArgGlyGluAspGlnSerMetLeuArgAspGluAlaThr 588
Db      1730 GCGAAATTAATGCGCTGTATCGTGTGAAAGATCAATCTGCTGCTGACGAGGCCACA 1789
Qy      589 Leu 589
Db      1790 CTG 1792

RESULT 4
AAA39722
ID AAA39722 standard; DNA; 1203 BP.
AC AAA39722;
XX
XX
XX 21-SEP-2000 (first entry)
DT
XX
XX Pseudomonas sp. HR199 fcs-delta DNA.
DE
XX
XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
KW vanillinic acid; flavouring; ferulic acid-CoA-synthetase; fcs; fcs-delta;
KW de.
XX
XX Pseudomonas sp.
XX
XX DE19850242-Al.
XX
XX PD 04-MAY-2000.
XX
XX PF 31-OCT-1998; 98DE-01050242.
XX
XX PR 31-OCT-1998; 98DE-01050242.
XX
XX PA (HAAR ) HAARMANN & REIMER GMBH.
XX
XX PI Rabenhorst J, Steinhuechel A, Priefert H, Overhage J;
XX
XX DR WP1; 2000-340642/30.
XX
XX DR P-PSDB; AA167972.
XX
XX PT Transformed organisms, used for production of vanillin and other
XX methoxyphenols, have altered catabolism of eugenol or ferulic acid.
XX
XX PS Claim 8; Fig 21; 80pp; German.
XX
XX CC This invention describes novel transformed and/or mutated uni- or multi-
XX cellular organisms (A) in which enzymes (I) involved in catabolism of
XX eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that
CC

```

```

QY 372 ProGlyCysGluValIysLeuValProValAspGlyIleuGluGlyArgPheHisGly 391
Db 1158 -----GACTGGCTGCTATTGGCGCAAGTCCGCGGG 1187
QY 392 ProHisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAlaGlnAlaPheArgIu 411
Db 1188 CAGGATCTCTGTCATCTAC---CTTGCTCTCCGCAAGAAATATCCATCAG--- 1238
QY 412 GluGlyTyrTrpCysSerGlyAspAlaIleLeuValLeuAlaAspProAlaAspPro 429
Db 1239 -----GCTGATGCATGCGCGCTGCATACGTTGATCCGGCTACC 1280
QY 430 -----GlnIys 431
Db 1281 TGGCCATTGCACACCAAGCAAAATCGCATGCAGCAGCAGCTACTCGATGGAAGCC 1340
QY 432 GlyLeuMetPheAspGlyArgIleAlaGluAspPheIysLeuSerSerGlyValAlaPheVal 451
Db 1341 GGTCTTGTGCATCAGATGATCTGCAGCAAGAGCATCAG-----GGGCTCGGCCCA 1391
QY 452 SerValGlyProLeuArgThrArgAlaValIleuGluGlyIleSerTyrValIleuAspVal 471
Db 1392 GCCGAACTGTTCCGCAAGCTCAAGCGCGCATGCGCGCAG---GATCTC 1442
QY 472 ValValAlaAlaIleProAspArgGluCysLeuGlyIleuLeuVal-----PhePro 487
Db 1443 GTCTGTAACCCATGCGCATGCTCTGCGCAATATGCTGTGAAATAGCGCCCTTTTCT 1502
QY 488 ArgLeuLeuAspCysArgAlaLeuSerGlyIleuGlyIysGluAlaSerAspAlaGluVal 1507
Db 1503 GGAATTCACTGCTGTGCGCGCTG---GCTGTGCGCAAGCCGCTATCAGCATACAGCTTG 1559
QY 508 LeuAlaSerGluProValAlaGalaIleTrpPheAlaAspTrpLeuIysArg-----Leu 524
Db 1560 GCTACCCGATGATATTCGTCGAAGAGCTTGCGCGCAATGGGCTGACCGCTTCTCTGTCCTT 1619
QY 525 AsnArgGluAlaThrArgAlaAsnAlaSerArgIleMetTrpValGlyLeuLeuAsp-ThrPr 544
Db 1620 TACGGTATCCCGCTCCCGCATTCGCAAGCGCATCGCTTCTATGCGCTTGTGACGAGTTTC 1679
QY 544 oProSerIleAspIysGlyIuValAlaThrAspIysGlySerIleAsnGlnArgAlaValIle 564
Db 1680 TTCTGAGCGGAGCTCTGGGCTTGAATATGCCAAGC-----GAGCCCTGTTT 1733
QY 564 uGlnTrpArgSerAlaIysValAspAlaLeuTyrArgGlyGluAspGlnSerMetLeuAr 584
Db 1734 GCAATGCGCGCTCGCGCAAGATGATGCGTGTATCGTGTGAAGATCAATCCATGCTGCG 1793
QY 584 GAspGluAlaThrLeu 589
Db 1794 TGACGAGCGCACACTG 1809

```

RESULT 3
AAA39721 standard; DNA; 2171 BP.

```

ID AAA39721 standard; DNA; 2171 BP.
AC AAA39721;
XX
XX 21-SEP-2000 (first entry)
XX
XX Pseudomonas sp. HR199 fcs-omega-Gm DNA.
XX
XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
XX vanillic acid; flavouring; ferulic acid-CoA-synthetase; fcs;
XX fcs-omega-Gm; ds.
XX
XX Pseudomonas sp.
XX
XX DE19850242-A1.
XX
XX 04-MAY-2000.
XX

```

```

PF 31-OCT-1998; 98DE-01050242.
XX
XX 31-OCT-1998; 98DE-01050242.
XX
XX (HAAR) HAARMANN & REIMER GMBH.
XX
XX Radenhorst J, Steinbuechel A, Priefert H, Overhage J;
XX P-PSDB; AA187971.
XX
XX WPI; 2000-340642/30.
XX
XX Transformed organisms, used for production of vanillin and other
XX methoxyphenols, have altered catabolism of eugenol or ferulic acid.
XX
XX Claim 8, Fig 2h; 80pp; German.
XX
XX This invention describes novel transformed and/or mutated uni- or multi-
XX cellular organisms (A) in which enzymes (I) involved in catabolism of
XX eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that
XX the organism accumulates the intermediates coniferyl alcohol (CAI),
XX coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
XX INDEPENDENT CLAIMS are also included for the following: (a) gene
XX structures in which sequences that encode one or more of the enzymes CAI-
XX dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-
XX aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase have been
XX altered and/or inactivated; (b) vector containing at least one gene
XX structure of (a); (c) biotechnical production of organic compounds
XX (particularly alcohols, aldehydes or organic acids) using (A); and (d)
XX production method of (A). (A) are used to prepare organic compounds,
XX preferably alcohols, aldehydes and organic acids, especially CAI, CA2,
XX FA, V and/or VA, most particularly V, a well-known natural flavouring.
XX The inexpensive starting material eugenol can be converted to V in a
XX single step. This sequence encodes the Pseudomonas sp. HR199 ferulic acid
XX -CoA-synthetase, fcs-omega-Gm which is described in the method of the
XX invention
XX
XX Sequence 2171 BP; 443 A; 591 C; 616 G; 521 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No. 1,818-59 Length: 2171
XX Score: 766.50 Matches: 227
XX Percent Similarity: 42.75% Conservative: 47
XX Best Local Similarity: 35.41% Mismatches: 158
XX Query Match: 24.87% Indels: 210
XX DB: 3 Gaps: 21
XX
XX US-09-750-986D-30 (1-589) x AAA39721 (1-2171)
QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyArgIleuGluIuArgLeuGluHis 20
Db 342 ATCCGTTCTCTGAGGCGCTTCTTCCCTTCCCGGCGCAATTTTAAAGGCTCGAGCAT 401
QY 21 TrpAlaIysThrArgProGluGlnThrCysValAlaIleArgAlaIleAsnGlyIuTrp 40
Db 402 TGGGCTAAGACCCGCTCAGAACAAACCTGCGTTGTCGACAGGCGGCAATGGGGAATGG 461
QY 41 ArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeu 60
Db 462 CGTCGTATCAGCTACCGGAAATGTTCCACACGTCGCGGCATTCGCACAGACTTGCTT 521
QY 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 522 CTTTACGCACTACCGGCAAGCGTCCGCTGCTTATGCTCTCTGGAATACCTCGAAACAT 581
QY 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCysAspProAla 100
Db 582 CTTCAAGCTGCAATTTGGGCTATGTATGCGGCGATTCCTATGCGCGGTGCTCTGCT 641
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaIysValAspArgHisIleValGlyLeuGluGlnPro 120
Db 642 TATTCACTGCTGCGCAAGATTGGCGAAGCTGCGTCAATGCTATGAGTCTTTCGCAACCG 701
QY 121 GlyLeuValPheAlaIleAspAlaIleProPheGlnArgAlaIleGluThrIleLeuPro 140

```

Db	1870	CTG	1872	
RESULT 2				
ID	AAA39720	standard; DNA; 2188 BP.		
XX	AAA39720;			
AC	21-SEP-2000	(first entry)		
XX				
DE	Pseudomonas sp. HR199	fccs-omega-Km DNA.		
XX				
KM	Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;			
KM	vanillic acid; Flavouring; ferulic acid-CoA-synthetase; fcs;			
KM	fccs-omega-Km; ds.			
XX				
XX	Pseudomonas sp.			
XX				
PN	DE19850242-Al.			
PD	04-MAY-2000.			
XX				
PF	31-OCT-1998;	98DE-01050242.		
PR	31-OCT-1998;	98DE-01050242.		
XX				
PA	(HAAR) HAARMANN & REIMER GMBH.			
XX				
PI	Rabenhorst J, Steinhuechel A, Priefert H, Overhage J;			
DR	WPI; 2000-340642/30.			
XX	P-PSDB; AAY87970.			
PT	Transformed organisms, used for production of vanillin and other			
XX	methoxyphenols, have altered catabolism of eugenol or ferulic acid.			
XX				
XX	Claim 8; Fig 2g; 80pp; German.			
CC	This invention describes novel transformed and/or mutated uni- or multi-			
CC	cellular organisms (A) in which enzymes (I) involved in catabolism of			
CC	eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that			
CC	the organism accumulates the intermediates coniferyl alcohol (CA1),			
CC	coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).			
CC	INDEPENDENT CLAIMS are also included for the following: (a) gene			
CC	structures in which sequences that encode one or more of the enzymes CA1-			
CC	dehydrogenase, CA2-dehydrogenase, PA-CoA-synthetase, enoyl-CoA hydratase			
CC	aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase have been			
CC	altered and/or inactivated; (b) vector containing at least one gene			
CC	structure of (a); (c) biotechnical production of organic compounds			
CC	[particularly alcohols, aldehydes or organic acids] using (A); and (d)			
CC	production method of (A). (A) are used to prepare organic compounds,			
CC	preferably alcohols, aldehydes and organic acids, especially CA1, CA2,			
CC	FA, V and/or VA, most particularly V, a well-known natural flavouring.			
CC	The inexpensive starting material eugenol can be converted to V in a			
CC	single step. This sequence encodes the Pseudomonas sp. HR199 ferulic acid			
CC	-CoA-synthetase, fcs-omega-Km which is described in the method of the			
CC	invention			
XX				
SO	Sequence 2188 BP; 408 A; 607 C; 670 G; 503 T; 0 U; 0 Other;			
Alignment Scores:				
Pred. No.:	2,04e-62	Length:	2188	
Score:	799.00	Matches:	234	
Percent Similarity:	45.53%	Conservative:	51	
Best Local Similarity:	37.38%	Mismatches:	168	
Query Match:	25.92%	Indels:	173	
BD:	3	Gaps:	20	
US-09-750-986D-30 (1-589) x AAA39720 (1-2188)				
1	1 MetAArgSerLeuGlnAlaLeuLeuPProPheProGlyValGllLeuGlnAArgLeuGlnHis 20			

Db	342	ATGGGTTCTCTCGAAGGGCGCTTCTTCCTCCCGGGGTGCAATTTCTTAGACGCTCGACAT	401
OY	21	TTpAlaLySThRaGrProGluGlnThCyVaLaLaAlaArgAlaAlaSerGlyLeuTrp	40
Db	402	TGGGCTAAAGCCCGTCCAGAACAAACCTGGCTGCTCCAGGGGGCAAAATGGGGAATGG	461
OY	41	ArgArgIleSerTYraIaGluMetPheHisAsnValArgAlaIleIaGlnSerLeuLeu	60
Db	462	CGTGGAATCACTACGGGGAATGTTCCACAACCTCCGCCATTCGACAGAGCTTGCTT	521
OY	61	ProTYrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGlnHis	80
Db	522	CCTTAACGACATATGGGACAGCGCTCGGCTGTTATCGCTCTGGAAATGACCTGAACAT	581
OY	81	LeuGlnLeuAlaPheGlyAlaMetTYraIaGlyIleProTYrCySProValSerProAla	100
Db	582	CTTGAGCTGGCATTTGGGGCTATGTATAGCCGGCGATTCCTATTCGCCGGTGTCTTCGCT	641
OY	101	TYrSerLeuLeuSerGlnAspLeuAlaIleValLeuArgHisIleValIleGlyLeuGlnPro	120
Db	642	TATTCACGCTGCTGCAGAAATTTGGCAGAGCTGGCATTCGTAAGATCTTCTGCAACG	701
OY	121	GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro	140
Db	702	GGACTGGCTTTGGTGGCGATGCAGACACTTTCAGGGG-----	740
OY	141	AspAspValProAlaIlePheThrArgGlyLeuLeuAlaGlyArgArgThrValSerPhe	160
Db	740	-----	740
OY	161	AspSerLeuLeuGlnProGlyIleGlyIleAspAsnAlaPheAlaAlaThrGly	180
Db	740	-----	740
OY	181	ProAspThrIleAlaIleAlaPheLeuPheThrSerGlySerThrIleLeuProValAlaVal	200
Db	741	-----ACAGCAAGCAACCGGAATTTGCCAGCTGGGGGG	773
OY	201	ProThrArgThrGlnArgMetLeuCyValAsnGlnIleMetLeuGlnThrPheProVal	220
Db	774	CCCTCTGGTAAG-----	785
OY	221	PheGlyGlnGluPProProValLeuValAspTrp-----LeuPro-----	233
Db	786	GTTTGGGAG-----CCCTGCMAAGTAACTGAGTGGCTTTCTTCCGCCAGAGATCTGA	839
OY	234	TrpAsnHisThrPheGlyGlySerHisAsnIleGlyIleValLeuTYrAsnGlyIleThr	253
Db	840	TGGCGCAGGGGATCAAGATCTGATCAAGACAGAGA-----TGAAGATCG	884
OY	254	TYrTYrLeu-----AspAspGlyLeuAspProThrAlaGlnGlyPheAlaGluThrLeuArg	271
Db	885	TTTGCATGATTTGAACAAGATGA-----	908
OY	272	AsnLeuSerGluIleSerProThrAlaTYrGlyLeuThrValProIleGlyTrpGlnGluLeu	291
Db	909	---TTGACGCGAGGTCTCCGGCCGCTGGGTGAGAGCTATTCCGCTATGACTGGGCA	965
OY	292	ValGlyAlaLeuGlnArgAspSerThrLeuArgGlyArgPhePheAlaArgMetLeuLeu	311
Db	966	CAACAGACATCGGCTGCTGATGCCGCCGCTGTCGGGCTGTAGAGCCAGAGGGCGCCG	1025
OY	312	PhePhePheAlaAlaAlaGlyLeuSerGlnGlyIleTrpAspArgLeuAspArgValAla	331
Db	1026	GTTCCTTTTTCAGACGCAACCTGTCCGGTCCCTGAATGAATCGACAGACAGGCGCG	1085
OY	332	GluGlnHisCysGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAla	351
Db	1086	---CGGCTATCGCTGGCCGACGAGGCGCTT-----	1115
OY	352	ProSerCyThrPheThrThrGlyProLeuSerMetAlaGlyTYrIleGlyLeuProAla	371
Db	1116	---CCTTGGCAGCTGTCTCGACGTTGATCTGAACCGGGAGAG-----	1157

PT New *Acinetobacter baumannii* proteins and nucleic acids, useful as reagents
 PT for diagnosing a bacterial disease, as components of antibacterial
 PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
 PT plants.

XX Example; SEQ ID NO 2393; 328bp; English.

CC The invention relates to isolated *Acinetobacter baumannii* nucleic acids.
 CC The *A. baumannii* nucleic acids and polypeptides are useful as reagents
 CC for diagnosing a bacterial disease, as components of antibacterial
 CC vaccines, as targets for antibacterial drugs, to detect the presence of
 CC *A. baumannii* and other *Acinetobacter* species in a sample, in screening
 CC compounds for the ability to interfere with the *A. baumannii* life cycle
 CC or to inhibit *A. baumannii* infection, and as biocontrol agents for
 CC plants. The present sequence represents DNA encoding an *A. baumannii*
 CC protein.

CC Sequence 1902 BP; 566 A; 420 C; 392 G; 524 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4,74e-140	Length:	1902
Score:	1652.50	Matches:	310
Percent Similarity:	70.57%	Conservative:	100
Best Local Similarity:	53.36%	Mismatches:	168
Query Match:	53.62%	Indels:	3
DB:	9	Gaps:	3

US-09-750-986D-30 (1-589) x ADB31106 (1-1902)

QY 5 GUAUAlaLeuLeuProPheProGlyArgGileuGluArgLeuGluHisTrpAlaLeuThr 24
 DB 133 GAGCATTAAACCGTATCCCAAAAAGTACAGATGCTTAATTCATTTCACAAACC 192
 QY 25 ArgProGluGlnThrCysValAlaAlaArgAlaAlaGluGluTrpArgGlySer 44
 DB 193 AAACCCGACCATATTTTTCAGCAAAAAGCAATGCTCAAGCGCATGGTCAAACTGAGT 252
 QY 45 TTAAlaGluMetPheHisAsnValArgAlaAlaGluGlnSerLeuLeuProTyArgGlyLeu 64
 DB 253 TATGCGAAGATCTTACCAACCGCATGCGACATGCTCAAGCTTTCAGCTCTGTAATTTA 312
 QY 65 SerAlaGluArgProLeuLeuLeuValSerGlyAsnAspLeuGluHisLeuAla 84
 DB 313 AGCCAAAGAAAGACTCTGCTGATTTTAAGTGTAATATCTCGAACCTTTAAACATGCTCT 372
 QY 85 PheGlyAlaMetCysValArgGlyLeuProTyArgProValSerProAlaTySerLeuLeu 104
 DB 373 ATGGCTGCCATGCTGCGACGCGTCCCTTCTCGGCTATTTCCCTGCTACTCTGATTT 432
 QY 105 SerGluAspLeuAlaValSerLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
 DB 433 TCTCAAGACTTGGCAAACTCAAAACATGTGTGTAAGTGTCTCAACCTGGTATGTTAT 492
 QY 125 AlaAlaAspAlaAlaProPheGlnArgAlaAlaGluThrIleLeuProAspValPro 144
 DB 493 GCCAAGGATGCAAGCTTTTGGCAAAAGCATTCAGCATGTAATTCAGCTGATATTGAA 552
 QY 145 AlaIlePheThrArgGlyLeuLeuAlaGlyArgArgThrValSerPheAspSerLeuLeu 164
 DB 553 GTGGTGCCTAATTAAGGAATAGTGGCGATCAGATCTGCAAGTCTTTTCATGCTGTTA 612
 QY 165 GluGluProGlyLeuIleGluAlaAspAsnAlaPheAlaAlaThrGlyProAspThrIle 184
 DB 613 GATACACCACTTTCA--AATGTTCAAGAGTTTATCAAACTTGATGAAGAAACAGATT 669
 QY 185 AlaIlePheLeuPheThrSerGlySerThrIleLeuProValAlaAlaProThrThrGln 204
 DB 670 GCCAAATTTCTGTTTACATCAGGTCAACTAAATTAATTAAGGTGTAACGACCAACAT 729
 QY 205 ArgMetLeuCysAlaGluGlnGlnMetLeuLeuGlnThrPheProValPheGlyGluGlu 224
 DB 730 TTAATGTTGTGTATATCAGCAAAATGTTATTCAGACTTCCCTGATTTGAAGAAACA 789

QY 225 ProProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGlySerHisAsnIle 244
 DB 790 CCGCTGTCTTACTCACTGCGTGTCTTGGCACCAACATTTGGCGGACTCAATGTC 849
 QY 245 GlyIleValLeuTyArgGluGlyThrTyArgLeuAspAspGlyLeuProThrAlaGln 264
 DB 850 GGCATGCACTATTAACGGCGGTGATGATTAATTAATGATGATGCAAAACCGTTGACGA 909
 QY 265 GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyLeuThrVal 284
 DB 910 AAATTTGACCAAACTATTTGTAATCTCAAGAAATTTCTCAACTGTTTAATTAATGTTG 969
 QY 285 ProGlySerGlyTrpGluGluLeuValGlyAlaLeuGluArgAspSerThrLeuArgGlyArg 304
 DB 970 CCAAAAGTTGGGAAGAACTCACCGAAGCTTGAAAGAAATGAAGATTAAGATGAGCCG 1029
 QY 305 PhePheAlaArgMetLeuSerPhePhePheAlaAlaGlyLeuSerGlnGlyIleTrp 324
 DB 1030 TTTTGGCCAAAGTAAATTTTATTTCTTGGCCGCTGTCACCTTCAGAAAGCGGCTGG 1089
 QY 325 AspArgLeuAspArgValAlaGluGlnHisCysGlyGlyLeuArgIleArgMetLeuAlaGly 344
 DB 1090 AACAGACTGATTAATTAATGCTCAGCAACATGGCGAAGAAATAATCCGATTAAGCCGA 1149
 QY 345 LeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAla 364
 DB 1150 TTGGCATGACCAAACTGCTCTTGTGCTTTTACAACTGCGCCACCGGTAAATGGCT 1209
 QY 365 GlyTyTrpIleGlyLeuProAlaProGlyCysGluValAlaLeuValProValAspGlyLeu 384
 DB 1210 GCGTTATTTATGTTATACCTGCTCGGATGCGAAATTAAGTATTCATGTTGTTGACAAA 1269
 QY 385 LeuGluGlyArgPheHisGlyProHisValMetSerGlyTyTrpArg--AlaProGlu 403
 DB 1270 CTTGAGTTTTCGCTTGTGCGCAAAACGTCATGAAGGCTATTGGCGCTTAAAGCGGAG 1329
 QY 404 GluAsnAlaGlnAlaPheAspGluGluGlyTyTrpCysSerGlyAspAlaIleValLeu 423
 DB 1330 CAACAAAGTACTTTTGTATGATGATGAAGCTTTTTCATACAGGCGATGCCGTTTGA 1389
 QY 424 AlaAspProAlaAspProGluLeuGlyLeuMetPheAspGlyValArgIleAlaGluAspPhe 443
 DB 1390 GTGATGTCATGATCAACCAAGGCTTAATGTAAGAGCAATTCGCGAAGCTTTT 1449
 QY 444 LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
 DB 1450 AAACCTAAATACAGCACTTTTGGAAATGTCGACACTACGCAAAAGTGTATTATCAA 1509
 QY 464 GlyGlySerTyArgValLeuAspValValAlaAlaProAspArgGluCysLeuGlyLeu 483
 DB 1510 GGTAAATTTACTCAATCCAAAGTGTGTATTAATCTGCTTCAAACTGAATGCTATGCTTT 1569
 QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGly--LysGluAla 502
 DB 1570 CTGATTTTTCAAATTAATGAAGCTGTGCTCAATATGACAGGCTTAAGTGGCGCAATAT 1629
 QY 503 SerAspAlaGluValLeuAlaSerGluProValArgAlaThrPheAlaAspTrpLeuLys 522
 DB 1630 TCTGACGACAGATTAATGCAACCTTAAGTCCAAATGTTCCGCCAATTTTAAAGC 1689
 QY 523 ArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAsp 542
 DB 1690 ACCTTAAATTAAGATCGACTGCGACTCAAAATACGTCTCAATGCTTTATTAAGACC 1749
 QY 543 ThrProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGlnArgAla 562
 DB 1750 GAACTCACTCAGTTAATATCCCGCGAAGTGAACGATTAAGGCAACCTCAATCAAAACAG 1809
 QY 563 ValLeuGlnTrpArgSerAlaLysValAspAlaLeuTyArgGlyLeuAspGlnSerMet 582
 DB 1810 ATTAACAAAGCTGTGCGCTTAAATTAATGAAGCTTTATCAAAAGCACTGATTAACCG 1869
 QY 583 Leu 583

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 7, 2005, 23:50:05 ; Search time 669 Seconds
(without alignments)
4621.688 Million cell updates/sec

Title: US-09-750-986D-30
Perfect score: 3082
Sequence: 1 MRSLEALPPGRILRLERH.....KVDALYRGEDQSMRLDEATL 589

Scoring table: BLASTSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 413486 seqs, 2624710521 residues
Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xid
-Q=/cgr2.1/USPTO.spool_P/US09750966/runat_03022005_071743_18555/app_query.fasta_1.775
-DB=N-Geneseq_23Sep04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCMALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09750966 @CGN 1.1 708 @runat 03022005 071743 18555 -NCPU=6 -ICPU=3
-NO MMAP -IAREBORETRY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEF_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N-Geneseq_23Sep04:*

1: geneseqn1980s:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1652.5	53.6	1902	9	ADA31106 DNA encod
2	799	25.9	2188	3	AAA39720 Pseudomon
3	766.5	24.9	2171	3	AAA39721 Pseudomon
4	698.5	22.7	1203	3	AAA39722 Pseudomon
5	314.5	10.2	349980	6	ABO81844 Bifidobac
6	309	10.0	38494	10	ADB74275 Mycobacte

7	3021.5	9.8	35961	10	ADB74389 Adb74389 Mycobacte
8	301.5	9.8	37286	4	AAS59522 Propionib
9	301.5	9.8	37286	4	ACF64451 ACF64451 Phototrab
10	290	9.4	1803	10	ACF71666 Continuation (49 o
11	290	9.4	110000	10	ACF67367_48
12	284	9.2	2103	11	ADL72353 Modified
13	283	9.2	2103	12	ADN72468
14	283	9.2	110000	2	AAT42063_00
15	279	9.1	2103	7	ADL23371 Plant acy
16	279	9.1	2103	11	ADL72249
17	278.5	9.0	349980	6	ABO81849
18	278	9.0	2106	7	ADL23370 Plant acy
19	277	9.0	2097	6	ABZ14464 Arabidops
20	277	9.0	2097	6	ADG87806 A. thalia
21	277	9.0	2097	8	ADA67967 Arabidops
22	275	8.9	2257	11	ADL72248
23	268.5	8.7	349980	6	ABO81846 Bifidobac
24	268	8.7	2262	3	AAC42047 Arabidops
25	267	8.7	2154	8	ADA70529 Rice gene
26	266.5	8.6	2094	8	ADA69386 Rice gene
27	266	8.6	1845	10	ADP02609 Bacterial
28	265.5	8.6	85692	12	ADL39159 Streptomy
29	264.5	8.6	1639	2	AA32852 Mutant lu
30	263.5	8.5	1639	4	AAS00899 Beetle cd
31	263.5	8.5	1639	4	AAS00894 Beetle cd
32	263.5	8.5	1639	6	ABO81820 Photuris
33	263.5	8.5	1639	12	ADM32125 Mutant lu
34	263	8.5	3264	10	ABV75591 Novel hum
35	263	8.5	6162	12	ADN04241 Antipso
36	262.5	8.5	1639	2	AA32853 Mutant lu
37	262.5	8.5	1639	4	AAS00895 Beetle cd
38	262.5	8.5	1639	6	ABO81818 Photuris
39	262.5	8.5	1639	12	ADM32123 Mutant lu
40	262.5	8.5	12951	8	ACA43808 Prokaryot
41	259	8.4	110000	4	AA199682_24 Continuation (25 o
42	259	8.4	110000	4	AA199683_24 Continuation (25 o
43	258	8.4	1639	4	AAS00898 Beetle cd
44	258	8.4	1639	6	ABO81819 Photuris
45	258	8.4	1639	12	ADM32124 Mutant lu

ALIGNMENTS

RESULT 1	
ID	ADA31106 standard; DNA; 1902 BP.
XX	ADA31106;
AC	ADA31106;
XX	20-NOV-2003 (first entry)
DT	
XX	DNA encoding Acinetobacter baumannii protein #2393.
DE	de; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
XX	kw vaccine; plant biocontrol agent.
KW	
XX	OS Acinetobacter baumannii.
XX	
PN	US6562958-BI.
XX	
PD	13-MAY-2003.
XX	
PF	04-JUN-1999; 99US-00328352.
XX	
PR	09-JUN-1998; 98US-0088701P.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Bretton G. Bush D;
XX	
DR	WPI; 2003-576092/54.
XX	P-PDB; ADA35232.
XX	

This Page Blank (uspto)

```
Db 170 FT-----LIYTSGETGEPKWMIDYONIAA--QVYLDQRLLTTEQDISISFLPLS 218
Qy 236 HTEGSHNIGIVLYNGSTVYLDGKPTAQGFATLRNLSEISPTAYLTPKWEELVGA- 294
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 219 HIFERAMSYVVMHTGALNVLTDTNPFREA-----MLDVRPTVMGSVRLYEKIHSAI 271
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 295 LERDSTLRERFFAPAMKLFPPAAGLSQGIWDR-----LDRVA---EQ 333
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 272 LEKVS--RAPFWR--IIFQWAIKRGERRMSROJLNDKSCPFSLCGEKLADRLVINKLRQ 327
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 334 HCGERTRMMA-----GLGWTETAPSGCTFTTGPLSMAGYIGLP 370
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 LILGGRIRFLPVAGARLEDSIMRFFLSVGLNIKYGYGTETCATVSCWEERXNYILGSGITP 387
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 371 APGCEVKLYVVDGKLEGRFPHGPMVMSGYWRAPENQAQFDEGYCSGDAIKLADPADPQ 430
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 388 LPGVDVR-ISTAGEIVR--GPIVMKGYNRPEDTVNHFTEDEGWLRTGDAGL---DET 440
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 431 KGLMPDGRITAEDFKLSGCVFVSVPPLKTRAVLEGGSYLDVVVAAPDRECLGLIVPPRL 490
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 GNLFIITERLKDLMKTSTGKYIA--POLIEGTLGQDRFIEHIAVIADARQFVSALIYP--- 495
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 491 DCRALSGIGKE-----ASDAEVLASEPVRAMPADWLKRLNREATGNASRLMWGLDTPP 545
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 496 SYEALEDYAKSINLHYQNRLLELNHNOVIGLFEORLKELOKN-IAKFHQVRRFTLLPESF 554
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 546 SIDKGEVTDKGSINORAVLQWRSKVDALYR 576
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 555 SMEKEGELTPTLKLRRKVISQRYQSEIESMYR 585
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: February 7, 2005, 23:59:47
job time : 109 secs

```

Db 302 ----SDVVISYLPPLAHYERANOIMGVYG---VAVGFYQGDVFKLMD----- 342
Qy 271 RNUSEISPTAYLTPVPGWELVGLERDSTLREFFAPMKLFFPAAAGLSGCI----- 323
Db 343 -DPAVLRPTIFCSVPLRYNRIYDGI--TSAVKSSGVKKRLFEIAYNSKKQAIINRTPS 399
Qy 324 --WDRL--DRVAEOHGERIRMA-----GLGMEETAPSGT 355
Db 400 AFMDKLVFNKIKK-LGGRVRFMGSGASPLSPVMDFLRICFGCSVREGYMET--SCV 456
Qy 356 FTT---GPLSMAGYIGLPAPGCEVKLVPD-----GKLEGFPHGVMSGW 399
Db 457 ISAMDDG-NLSGHVSPNPACCEVKLVDPENMYTSDQPYPRGEICVR--GPILFKGY 513
Qy 400 RAPEOMAQAFDEGYGCSGDAIKLADPADPQKGLMPD-----RIADFKLSSGVFS 452
Db 514 KOBEOQREILDDGMLHTGDI-----GLMIPGRLKIIDRKNIIFKLAQGEYIA 562
Qy 453 VGFLR-----TRAVLEGGSY---VLDVVVAAPD-----RECLGLLV-PP 487
Db 563 PEKIEVVTYKCRFVSCGFIHGDSFNLSVAIVSDPEVKMDMAASGIKYEHGQLCNDP 622
Qy 488 RLID--CRALSGLGEA-----SDAEVLASEP-----VRAMPADW 520
Db 623 RVAKTVLAEMDDLGREAQURGEFEFAKAVTLVPEPFTLENGLLTPTFKIKRPAKAYFAEA 682
Qy 521 LKRLNE 527
Db 683 ISKRYAE 689

```

RESULT 14

```

US-10-410-031-130
; Sequence 130, Application US/10410031
; Publication No. US20040010817A1
; GENERAL INFORMATION:
; APPLICANT: Shockley, Jay M.
; APPLICANT: Schmutz, Judy
; APPLICANT: Browne, John A.
; TITLE OF INVENTION: Plant Acyl-CoA Synthetasees
; FILE REFERENCE: DOM-07654
; CURRENT APPLICATION NUMBER: US/10/410,031
; CURRENT FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 130
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-410-031-130

```

```

Query Match 9.2%; Score 283; DB 15; Length 700;
Best Local Similarity 22.5%; Pred. No. 8.8e-18;
Matches 150; Conservative 97; Mismatches 214; Indels 206; Gaps 35;

```

```

Db 302 ----SDVVISYLPPLAHYERANOIMGVYG---VAVGFYQGDVFKLMD----- 342
Qy 271 RNUSEISPTAYLTPVPGWELVGLERDSTLREFFAPMKLFFPAAAGLSGCI----- 323
Db 343 -DPAVLRPTIFCSVPLRYNRIYDGI--TSAVKSSGVKKRLFEIAYNSKKQAIINRTPS 399
Qy 324 --WDRL--DRVAEOHGERIRMA-----GLGMEETAPSGT 355
Db 400 AFMDKLVFNKIKK-LGGRVRFMGSGASPLSPVMDFLRICFGCSVREGYMET--SCV 456
Qy 356 FTT---GPLSMAGYIGLPAPGCEVKLVPD-----GKLEGFPHGVMSGW 399
Db 457 ISAMDDG-NLSGHVSPNPACCEVKLVDPENMYTSDQPYPRGEICVR--GPILFKGY 513
Qy 400 RAPEOMAQAFDEGYGCSGDAIKLADPADPQKGLMPD-----RIADFKLSSGVFS 452
Db 514 KOBEOQREILDDGMLHTGDI-----GLMIPGRLKIIDRKNIIFKLAQGEYIA 562
Qy 453 VGFLR-----TRAVLEGGSY---VLDVVVAAPD-----RECLGLLV-PP 487
Db 563 PEKIEVVTYKCRFVSCGFIHGDSFNLSVAIVSDPEVKMDMAASGIKYEHGQLCNDP 622
Qy 488 RLID--CRALSGLGEA-----SDAEVLASEP-----VRAMPADW 520
Db 623 RVAKTVLAEMDDLGREAQURGEFEFAKAVTLVPEPFTLENGLLTPTFKIKRPAKAYFAEA 682
Qy 521 LKRLNE 527
Db 683 ISKRYAE 689

```

RESULT 15

```

US-10-369-493-311
; Sequence 311, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 311
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-311

```

```

Query Match 9.1%; Score 281; DB 14; Length 585;
Best Local Similarity 22.2%; Pred. No. 1.1e-17;
Matches 140; Conservative 105; Mismatches 268; Indels 118; Gaps 21;

```

```

Qy 9 PFGRILIEREHAHAKRREBQTCVAARAAN-----GEMRISYAMFHNVAIAQSLLPYGL 64
Db 81 PBTGLHDFVHAVENTAENKYLSTRVSDGTIGETYSMMTYGEASERQAIISGLLFHGV 140
Qy 65 S-----AERPLILVSGNDLEHLQALFAGAMVAGIPYCPVSPAYSLISQDLAK--LRH 113
Db 141 NQDDCVGLYFIRNPEMLV---VDHCAAYSFVSV-----PLVDTGLPADAVKPVVNH 188
Qy 114 -----IVGLIQGLVPAADAAPFORAIETI--LPDDVPAIFTRGELAGRTVSFDSL 163
Db 189 ANIOAIFCVQPTNIIISFLAE--IPSIIRLIIVVGADEHLPSTL-PRG--TGVTIVSYOKL 244
Qy 164 LEOPGIEANARANGPDTIAKFLFTSGSTKLKPAVPTQRMICAN---QOMLOTFPV 220
Db 245 LSG--GSSSHAPSPKPEBDAIITCYTSGTGTGPKGVLLHGNLIANVASSVEAEFFP- 301
Qy 221 FGEEPPVLVMDLPWNH-----TFGSHNIGIYLVNGGTYYLDDGKPTAQGFATL 270

```

```

Qy 18 LEHWATREBQTCVAARAANGEMRISYAMFHNVAIAQSLLPYGLSAERPLIVSGN- 76
Db 1 LQOQVSKYEKIAFRQWSAS--EGLKMSQDQVDSKTSIARSLSLSEVGIQEKIGIPADS 59
Qy 77 -----DLEHLQALFAGAMVAGIPYCPVSPAYSLISQDLA-----KLRHVLGLQPG 121
Db 60 MAWSLSDLLIQLR-----AVTVPLVATSSRDOAFILNDAGVAILLFVGGQEQYD 109
Qy 122 LVFAADAAPFORAIETILPDV-----PAIFTRGELAGRTVSFDSLLEOPGIEADNA 175
Db 110 IALILALCPQNLNLVLDSEVDLKECPQALYISATLADRPVYQAELEDSRISAGDILHL 169
Qy 176 FAATGDTIAKFLFTSGSTKLKPAVPTQRMICANQOMLOTFPVFGEEPPVLVMDLPWN 235

```

```

QY 298 2STLREREPARKU-----FFFA----AG-----LSGG--IMWDLDRPAEHCGERIR 340
Db 295 DMOKNQGFPPYINQLKANKIIRSKREALGAGNVITVSGAALQPRLAR--FWAAQIP 351
QY 341 MMAAGLGMTETAPSCTF-----TGPLSMAGYIGLPAPGCWVKLVPPVGLKEGRF 389
Db 352 VMEGGGLTETSVIVANRPNPDETRISTYGPL-----LPGVEVKIAE-DGEILTR- 400
QY 390 HGPVMSGYWRAPAEONAOAFDEEGYCSGDAIKLADPADPDKGLMFDGRJAEDFKLSGV 449
Db 401 -GPHIMKGYYKRPDLTAEALDSEGWHHTDIGQFVE---NKYLKITTRKKEHFKTSGK 455
QY 450 FVSVGPLTRAVLEGGSYVLDVVAAAPRECTGLVFPPLDGRA--LSGLCKEASDAE 506
Db 456 YIA--POLVENFKESRYEQIEMVIGEGRKFPASIVPSFADVKSCEIKGI-PTYTNEE 512
QY 507 VLASFPVAAPFADWLKRLNREATGNASRIIMWGGLDTPPSIDKGEYTDKGSINQRAVLQW 566
Db 513 MVCAAQVIEKQKEPEDEMN-ENFPAOYERKFKETLMRPRTISGEMTPPLKLRKTIIMON 571
QY 567 RSAKVDALY 575
Db 572 CQGQIEKNY 580

```

```

RESULT 12
US-09-906-419-17
; Sequence 17, Application US/09906419
; Publication No. US20030037357A1
; GENERAL INFORMATION:
; APPLICANT: Schmuur, Jay
; APPLICANT: Schmuur, Judy
; APPLICANT: Browne, John
; TITLE OF INVENTION: Plant Acyl-CoA Synthetases
; FILE REFERENCE: DOW-04679
; CURRENT APPLICATION NUMBER: US/09/906,419
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 60/220,474
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-09-906-419-17

```

Query Match	9.2%;	Score 283;	DB 10;	length 700;
Best Local Similarity	22.5%;	Pred. No. 8.8e-18;		
Matches 150;	Conservative 97;	Mismatches 214;	Indels 206;	Gaps 35;

```

OY 9 FPPGGIIEERLEEMATREGEQTCVABAAN-----GEMRRISYEMEFNNVAIQSLIPYGI 64
Db 81 PEIGTLHNFPHAVETIYENKLTGTRVNSDGTIGETSMYTGEBASEROAQISGGLFHG 140
OY 65 S-----AERPLLIYSGNDLEHLOLAFGAMYGI PYCPVSPAYSLISODLAK--LRH 113
Db 141 NQGDCEVGLYFIRNPEMLV-----VDHACAYSPVSU-----PLVDTLGPDPAKRVVNH 188
OY 114 -----IVGLIQGLVFPADAPAPORAIETI--LPDDVPRAIFTRGELAGRRTVSFDSI 163
Db 189 ANLQAFICVPOTLNLITSLFLAE--ISIRIIVAVGADHPLSL-PRG--TGVTIVSYOKI 244
OY 164 LEOPGIEADNAPFAATGPDTIAKPLFTSGSTKLPRKAPPTQOMLCAN-----QQMLLTQFPV 220
Db 245 LSQ--GRSSLHFSPKPEDIATICTYISCTGTGPKVULLHGNLIINAVAGSSVEAEFFP- 301
OY 221 FGEBBPVLVDMLPMNH-----TFGSGHNIGIVLYNGGTYVLDDEKPTAQGEAETL 270
Db 302 ----SDVYISYLPPLHIYERANQINGVGG--VAVGFQSGVDFKMD----- 342
OY 271 RNLEISITAVIUYKGMVELGALERDSTLERFFARKKLFPFAAAGLSQI----- 323

```

[illegible]

```

RESULT 13
US-10-119-136-130
; Sequence 130. Application US/10119136
; Publication No. US20030097676A1
; GENERAL INFORMATION:
; APPLICANT: Shocke, Jay
; APPLICANT: Schmitt, Judy
; APPLICANT: Browne, John
; TITLE OF INVENTION: Plant Acyl-CoA Synthetases
; FILE REFERENCE: DOW-04695
; CURRENT APPLICATION NUMBER: US/10/119,136
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 09/906,419
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 60/220,474
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 130
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-119-136-130

```

Query Match	9.2%;	Score 283;	DB 14;	Length 700;
Best Local Similarity	22.5%;	Pred. No. 8.8e-18;		
Matches 150;	Conservative 97;	Mismatches 214;	Indels 206;	Gaps 35;

```

0Y 9 PEPGILRLHEHMAATREQCCTVAARAAN-----GEMRRTSYAEMFNVAIAKOSLIPYGI 64
Db 81 PEIGLHNPFAHVAETVAENKTLGTRVNSDGTIGESMWTYGEASAERAIISGLLFHCY 140
0Y 65 S-----ABRPILIVSGNDLEHLQIAFGAMYAGIPLYCPVSPAYSLISODLAK--LKH 113
Db 141 NOGDGCVGLYFINRPFWLV-----VDHACAAYSFVSU-----PLYDTGLGPDAVKRVVNH 188
0Y 114 -----IVGILPGLVPAADAAFPORAIEFT--LPDDVPAIFTRGELAGRRTVSPDSI 163
Db 189 ANLOAIFCVQPTNLNILLSFLAE--IPSILIVVAGADEHLFSL--PRG--TGWTIVSYOKL 244
0Y 164 LEQGGIIEADNAAPATGPDITAKFLPFTSGSTLCPRAVPTQMLCAN---QQMLLOTFPV 220
Db 245 LSO--GRSLHPFSPKPEDIATICTYTGTTCTPRGCVLTHGNLIANVAGSSVEAEFFP-- 301
0Y 221 FGEBPVLVLDMLPMNH-----TFGSHNIGIVLVNGSTYYLIDGKPTAOGFAETL 270

```

RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Ouriello B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,
 RA Hounmel K., Gordon J., Vaudin M., Tatchouk O., Bep A., Liu F.,
 RA Moliam C., Allinger M., Doughty D., Scott C., Lappas B.,
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.,
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58."
 RL Science 294:2323-2328(2001).
 CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
 CC family.
 CC EMBL: AEO08067; AAK87208.1; -
 DR InterPro: IPR000873; AMP-bind.
 DR Pfam: PF00501; AMP-binding. 1.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00455; AMP_BINDING. 1.
 SQ SEQUENCE 591 AA; 65535 MW; 9B371D0064CBE856 CRC64;
 Query Match 42.0%; Score 1294; DB 2; Length 591;
 Best Local Similarity 46.5%; Pred. No. 7.8e-86;
 Matches 266; Conservative 84; Mismatches 220; Indels 2; Gaps 2;
 QY 5 EALLPFGRIERLEHMAKTRPEQTCVAPAAANGEMRISYAEFHNVRVIAQSLPYGL 64
 DB 6 DPLGYPQKLNRLHMCRAPERTWADRGREPMRVSYAALDKIRRIQGLDHL 65
 QY 65 SARRPLIVSGNDLEHLQAFGMYAGIPYCPVSPAYSLISODLAKRHHVGLQPLVF 124
 DB 66 SVRRPLIVSENSIEHALMVAQHVIAASAIAPAYA-TSADTKLADIRGQITPMVF 124
 QY 125 AAAAARFORAIEITLPPDVPAIFTRGELAGR-RTVSFDSLLEQGGIEADNAFAATGPD 183
 DB 125 AEDATPPRRALGVEFDDGTPLVGLRLNLPEDRSNTFHETLLETBPTAVIDRAFAVGPDT 184
 QY 184 IAKFLFTSGSTKLKPKAVPTTORMLCANQMLLOTFPVFGSEPPVLDMLPMNHTFGSHN 243
 DB 185 VAKFLFTSGTTSKPAVITQORMLCNQEMIAADCYGFREPPVVDMAFMNHTAAGNKV 244
 QY 244 IGVLYNGGTYIYDDGKPTQAGFAETLRNLSISPTAYLYVPKWEELVGLALEDSTLRE 303
 DB 245 FNVLYNGGTYIYDRGKPSPAQIQGTLDNRDISPTVFNVPAGHEMLVQAMRKDEALCR 304
 QY 304 RFPARMKLFPPAAGLSQGIWDRLDRAVEQHCGERIRMAAGLGMTETAPSCFTTGLSM 363
 DB 305 SFPRDLKMLVYAGMAQHTWDALTELSMATVGAIVMGAGLSTERTAPPSLCTEPQDK 364
 QY 364 AGYIGLPAPGCEVYLVDGKLEGRFHGPHVMGMYRPAEQNAQAPDEEGYCSGDAIKL 423
 DB 365 PGNIGIPAGQVTMKLVFPDGRYELRLKGNPITPGYMNNGELTAAAPDEEGYRIGDTVKF 424
 QY 424 ADPADPQKGLMPDGRVIAEDFKLSGVFVSQPLTRAVLEGGSYVLDVVVAAPDEECLGL 483
 DB 425 AVADDPKRGYFPDGRMAENFKLOTGTWVAAGPLAQLVNMFAGLIRDAVITGEMRAELGA 484
 QY 484 LVPRRLDRCALSGLKEASDAEVLASEPVRAMPADMLKRLNREATGNASRIMVGLD 543
 DB 485 LVVPFIRPALRELVRGSHLSDAETIRHPSVRAQIVAKLSHQKQASGSASRVMRIIWMED 544
 QY 544 PSIDKGEVTDKGSINQRAVLQWMSAKVDALY 575
 DB 545 ALRFEKGEVTDKGSINQRAVLRLHRELVESELY 576
 RESULT 15
 ID Q8UF11 PRELIMINARY; PRT; 611 AA.
 AC Q8UF11
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Long-chain-fatty-acid-CoA-ligase.
 GN OrderedLocusNames=Atu1416;
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 OC NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Dupont;
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monke D.E., Kiteajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D., St.
 RA Chapman P., Clendenning J., Decherage G., Gillet W., Grant C.,
 RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.,
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58."
 RL Science 294:2317-2323(2001).
 CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
 CC family.
 CC EMBL: AEO09102; AAL42422.1; -
 DR PIR: AH2750; AH2750.
 DR PIR: G97531; G97531.
 DR GO: GO:0016874; F.ligase activity; IEA.
 DR GO: GO:0008152; P.metalloion; IEA.
 DR InterPro: IPR000873; AMP-bind.
 DR Pfam: PF00501; AMP-binding. 1.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00455; AMP_BINDING. 1.
 KW Complete proteome; Ligase.
 SQ SEQUENCE 611 AA; 67860 MW; D86FF0A396B058B0 CRC64;
 Query Match 42.0%; Score 1294; DB 2; Length 611;
 Best Local Similarity 46.5%; Pred. No. 8.1e-86;
 Matches 266; Conservative 84; Mismatches 220; Indels 2; Gaps 2;
 QY 5 EALLPFGRIERLEHMAKTRPEQTCVAPAAANGEMRISYAEFHNVRVIAQSLPYGL 64
 DB 26 DPLGYPQKLNRLHMCRAPERTWADRGREPMRVSYAALDKIRRIQGLDHL 85
 QY 65 SARRPLIVSGNDLEHLQAFGMYAGIPYCPVSPAYSLISODLAKRHHVGLQPLVF 124
 DB 86 SVRRPLIVSENSIEHALMVAQHVIAASAIATTPAYA-TSADTKLADIRGQITPMVF 144
 QY 125 AAAAARFORAIEITLPPDVPAIFTRGELAGR-RTVSFDSLLEQGGIEADNAFAATGPD 183
 DB 145 AEDATPPRRALGVEFDDGTPLVGLRLNLPEDRSNTFHETLLETBPTAVIDRAFAVGPDT 204
 QY 184 IAKFLFTSGSTKLKPKAVPTTORMLCANQMLLOTFPVFGSEPPVLDMLPMNHTFGSHN 243
 DB 205 VAKFLFTSGTTSKPAVITQORMLCNQEMIAADCYGFREPPVVDMAFMNHTAAGNKV 264
 QY 244 IGVLYNGGTYIYDDGKPTQAGFAETLRNLSISPTAYLYVPKWEELVGLALEDSTLRE 303
 DB 265 FNVLYNGGTYIYDRGKPSPAQIQGTLDNRDISPTVFNVPAGHEMLVQAMRKDEALCR 324
 QY 304 RFPARMKLFPPAAGLSQGIWDRLDRAVEQHCGERIRMAAGLGMTETAPSCFTTGLSM 363
 DB 325 SFPRDLKMLVYAGMAQHTWDALTELSMATVGAIVMGAGLSTERTAPPSLCTEPQDK 384
 QY 364 AGYIGLPAPGCEVYLVDGKLEGRFHGPHVMGMYRPAEQNAQAPDEEGYCSGDAIKL 423
 DB 385 PGNIGIPAGQVTMKLVFPDGRYELRLKGNPITPGYMNNGELTAAAPDEEGYRIGDTVKF 444
 QY 424 ADPADPQKGLMPDGRVIAEDFKLSGVFVSQPLTRAVLEGGSYVLDVVVAAPDEECLGL 483
 DB 445 AVADDPKRGYFPDGRMAENFKLOTGTWVAAGPLAQLVNMFAGLIRDAVITGEMRAELGA 504
 QY 484 LVPRRLDRCALSGLKEASDAEVLASEPVRAMPADMLKRLNREATGNASRIMVGLD 543
 DB 505 LVVPFIRPALRELVRGSHLSDAETIRHPSVRAQIVAKLSHQKQASGSASRVMRIIWMED 564

Qy 544 PSIDKGEVTDKGSINQRAVLQWRSACVDALY 575
:|||||:|
Db 565 ALRFEKGEVTDKGSINQRAVLHLHRELVESTLY 596

Search completed: February 7, 2005, 23:47:44
Job time : 147 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 7, 2005, 23:51:01 | Search time 5877 Seconds
(without alignments)
4739.430 Million cell updates/sec

Title: US-09-750-986D-30

Perfect score: 3082
Sequence: 1 MRSLEALPFGRIERLEH.....KYDALVRGDSQSLRDEATL 589

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 segs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+pin.model -DEV=xip
-O=/cgn2.1/USPTO.spool_p/US09750986/runat_03022005_071743_18565/app_query.faeta_1.775
-DB=GenEmbl -QFMT=fastp -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNIT8=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45
-OCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09750986@CN1_1_5600@runat_03022005_071743_18565 -NCPU=6 -ICPU=3
-NO_WMAP -LANG=QUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WAPN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.*
1: gb_da:.*
2: gb_hcg:.*
3: gb_in:.*
4: gb_om:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_pro:.*
11: gb_sre:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3082	100.0	1770	6	A92118 Sequence 29
2	3082	100.0	4166	1	AJ238746 Pseudomon
3	3082	100.0	32679	6	A92090 Sequence 1
4	2474	80.3	310029	1	AE016866 Pseudomon

5	2449	79.5	9477	1	PFL56325	AJ536325 Pseudomon
6	2349	76.2	301214	1	AB016786	AE016786 Pseudomon
7	2103	68.2	4407	1	PEU536324	AJ536324 Pseudomon
8	1940.5	63.0	189050	1	AL646077	AL646077 Ralstonia
9	1652.5	53.6	1902	6	AR319843	AR319843 Sequence
10	1621.5	52.6	45324	1	ACCECAOP	L05770 Acinetobact
11	1618.5	47.5	110000	1	CR543861_17	Continuation (18 o
12	1457	47.3	346879	1	BS572598	BX572598 Rhodosphe
13	1401	45.5	299950	1	AP005963	AP005963 Bradyrhiz
14	1294	42.0	10173	1	AE008067	AE008067 Agrobacte
15	1294	42.0	10173	1	AE009102	AE009102 Agrobacte
16	1210.5	39.3	299450	1	AP005938	AP005938 Bradyrhiz
17	1129	36.6	299950	1	AP005963	AP005963 Bradyrhiz
18	1002.5	32.5	10306	1	AE005712	AE005712 Caulobact
19	972	31.5	12606	1	AE005995	AE005995 Caulobact
20	799	25.9	2188	6	AX024541	AX024541 Sequence
21	766.5	24.9	2171	6	AX024537	AX024537 Sequence
22	743	24.1	7171	1	AF302797	AF302797 Pseudomon
23	698.5	22.7	1203	6	AX024534	AX024534 Sequence
24	337	10.9	301457	1	AE016924	AE016924 Chromobac
25	335.5	10.9	299925	1	AP005043	AP005043 Streptomy
26	325	10.5	10395	1	AE012877	AE012877 Chlorobiu
27	319.5	10.4	310550	1	SC0939113	AL939113 Streptomy
28	314.5	10.2	10502	1	AE014645	AE014645 Bifidobac
29	314.5	10.2	349980	6	AX492782	AX492782 Sequence
30	314.5	10.2	349980	6	AX553949	AX553949 Sequence
31	310.5	10.1	2109	6	BX936010	BX936010 Gallus ga
32	309	10.0	38494	6	AR345349	AR345349 Sequence
33	309	10.0	38503	1	MSG81912CS	L01536 M. lepreae g
34	302.5	9.8	35961	1	MLUJ5181	U15181 Mycobacteri
35	302.5	9.8	35961	6	AR345370	AR345370 Sequence
36	302.5	9.8	348450	1	MLBPRTN4	AL589920 Mycobacte
37	301.5	9.8	37286	6	CQ363734	CQ363734 Sequence
38	301.5	9.8	110000	1	AE017283_24	Continuation (25 o
39	294.5	9.6	11198	1	AE004318	AE004318 Vibrio ch
40	294.5	9.6	248950	1	SC0939120	AL939120 Streptomy
41	290	9.4	349944	1	BX571871	BX571871 Photocorhab
42	289	9.4	10029	1	AE013966	AE013966 Yersinia
43	289	9.4	201050	1	AJ414143	AJ414143 Yersinia
44	289	9.4	293371	1	AE017141	AE017141 Yersinia
45	289	9.4	300900	1	AP005939	AP005939 Bradyrhiz

ALIGNMENTS

RESULT 1	A92118	Sequence 29 from Patent EP0845532.	1770 bp	DNA	linear	PAT 22-JAN-2000
LOCUS	A92118					
DEFINITION	A92118					
ACCESSION	A92118					
VERSION	A92118.1	GI:6740914				
KEYWORDS						
SOURCE	unidentified					
ORGANISM	unclassified					
REFERENCE	1 (bases 1 to 1770)					
AUTHORS	Priefer, H.D. and Rabenhorst, J.D.					
TITLE	Enzymes for the synthesis of coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin, vanillic acid and their applications					
JOURNAL	Patent: EP 0845532-A 29 (03-JUN-1998);					
HAARMANN & REIMER GMBH (DE)						
FEATURES	Location/Qualifiers					
source	1..1770	/organism="unclassified"				
gene	1..1770	/mol_type="unassigned DNA"				
CDS	1..1770	/db_xref="taxon:32644"				
	1..1770	/gene="PCS"				
	1..1770	/gene="PCS"				
	1..1770	/codon_start=1				
	1..1770	/product="FERULASACURE-COA-SYNTHETASE"				
	1..1770	/protein_id="CAB69489.1"				


```

/db xref="GI:6740915"
/translation="MRSLEBALPFGRIILERLEHMAKTRPEQTCVVARANBEMBRIS
VEMFNVAIAIOSLIPGLSARPLIYSGNDLEHLQIAFGAMVAYICYCPSPYS
ILSOLAKIRHTVGLQPGIVRAADAPROAIRLITLPODVAPLFRGLARRTYSP
DSLLEQPGITEADNAPATGPDITAKFLTSGSTKLPKAVPTTORMLCANQMLDTF
PVFGEPPVLVWMLPMNHTFGGSHNIGIYLVNGTYLDDGKPTAGCFEETLRNISEI
SEPTAYLTVPKGMELVGLERDSTLEREPARMKLPFPAAGLSQGIWRLDRVVEOH
CEERIIMAGLMTETAPSCFTTGTPLSAAGYIGLPAFCGEVLVVDGKLEFGRFHP
HVMGWRAPBEONAOAFDEEGYCSGDALKADPADPOKGYMDGR IAEFDKLSGVF
VSGPLRMPAVLEGGSYVLDVVVAAPRECIQLIYVPRLLIDCRALSGLKEASDEVL
ASEPVRAPADMLKRLNREATGNASRIIMVGLIDTPPSIDKEBVTDKSINQRAVLQW
RAKVDALYRGDSMLRDEATL"

```

ORIGIN

Alignment Scores:

Pred. No.:	9,38e-211	Length:	1770
Score:	3082.00	Matches:	589
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-750-986d-30 (1-589) x A92118 (1-1770)

```

QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyArgIleLeuGluuArgLeuGluHs 20
DB 1 ATGCGTCTCTCGAGCGCTTCTTCCCTCCCGGCGTGAATTCCTTGAGCGTCTCGAGCAT 60
QY 21 TTPAlaIyGlyThrArgProGluGlnThrCyValAlaAlaArgAlaAlaIleGluThr 40
DB 61 TGGCGTAAACCCCGTCACAAACCAACCTCGTTCGCGACGAGCGGCAATGGGGAATGG 120
QY 41 ATGArgIleSerTyrAlaGluMetPheHsAsnValArgAlaIleAlaGlnSerLeuLeu 60
DB 121 CCGTCGATACGCTACCGGAAATGTTCCACACAGTCCGCGCATCCACAGAGCTTGCTT 180
QY 61 ProTyrGlyLeuSerAlaGluuArgProLeuLeuIleValSerGlyAsnAspLeuGluHs 80
DB 181 CCTTACGGAGCTATCGCGACAGCGCTCGCTTATCGTCTCGGAATGACCTCGGAACAT 240
QY 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCyAspProValSerProAla 100
DB 241 CTTTACGCTGCGATTGGGGCTATGATGGGGCATTCCTATTTGCCGGGTCTCTGCT 300
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaIySLeuArgHsIleValGlyLeuLeuGlnPro 120
DB 301 TATTACCTGCTGCGAAGATTGGGGAAGCTGCTGCACATCGTAGCTTCTGCAACCG 360
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
DB 361 GGAAGTGTCTTGTCTCCGATGACGACCTTTCACAGCGCAATTGAGACCATTTGCGCG 420
QY 141 AspAspValProAlaIlePheThrArgGlyLeuLeuAlaGlyArgArgThrValSerPhe 160
DB 421 GACGAGTGCCTCCGCAATCTTCACTCGAGGAAATTGGCGGGCGGCGCAACGGATTTT 480
QY 161 AspSerLeuLeuGluGlnProGlyIyIleGluAlaAspAsnAlaPheAlaIleThrGly 180
DB 481 GACAGCTCTCGAGAGAGCTGTGGGATTGAGGCAAGATATGCCCTTTGCGGAACTGGC 540
QY 181 ProAspThrIleAlaIySLeuPheLeuPheThrSerGlySerThrIySLeuProIySAlaVal 200
DB 541 CCCGATACGATGGCCAAATCTTGTTCACCTTCTGAGCTTACCAAACTGCTTGAAGCGGTG 600
QY 201 ProThrThrGlnArgMetLeuCyAlaIleGlnGlnMetLeuLeuGlnThrPheProVal 220
DB 601 CCGACTACTCAGGGAATGCTGTGGCCATCAGACATGCTTCTGAAACTTTCCTGGGTT 660
QY 221 PheGlyGlyGluProProValLeuValAspTyrLeuProTyrPheHsIleThrPheGlyGly 240
DB 661 TTGGGGAAGAGCGCGGCTGCTGAGTGTGATTCCTGGAACACACACTTTCGCGCGC 720
QY 241 SerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspAspGlyIyS 260

```

```

DB 721 ACCCAACAATCGGATCGTGTGTGACAAAGCGGCGCATGACTTACCAACCGGTAA 780
QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
DB 781 CCACCGGCCCAAGGGTCTCCCGAGAGCTTCCCACTTGAAGGAATCTCTCCCATGGG 840
QY 281 TyrLeuThrValProIySgIyTyrGluGluLeuValGlyAlaLeuGluuArgAspSerThr 300
DB 841 TACCTCAGCTGCTCCGAAGAGCTGGAGGAATAGTGCGTGCCCTTGAGGAGACAGTACC 900
QY 301 LeuArgGluArgPhePheAlaArgMetIySLeuPhePhePheAlaAlaIleGlyLeuSer 320
DB 901 CTGCGCGAAACGCTTCTTCCCTCGCATGAAGCGTCTTCTTCCGCGCGCTGGGTGTG 960
QY 321 GlnGlyIleTyrAspArgLeuAspArgValAlaGluGlnHsCyGsgIyGluArgIleArg 340
DB 961 CAAGGATCTGGGATCGTGTGACCGGGTGGCTGAACAGCACTGTGTGAGGCGCATTTCC 1020
QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCyThrPheThrThrGlyPro 360
DB 1021 ATGATGGCGGCTGTGGGATGACGAGGACTGCTCTTCTGCACTTTACACCGGACCG 1080
QY 361 LeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCyGsgIyValIySLeuValPro 380
DB 1081 CTGTGATGCTGCTGTACATTTGGGCTGCCAGCGCTGGCTGGAGGTCAAGCTCGTTCCG 1140
QY 381 ValAspGlyIySLeuGluGlyArgPheHsIleGlyProHsIleValMetSerGlyTyrTyrPArg 400
DB 1141 GTGATGGGAAATTTGGAAGGGGCTTTCCATGTTCCGACCTATGACGGGCTACTGGCCT 1200
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCySserGlyAspAla 420
DB 1201 GCTCCGAAACAATGGCCCAAGCGTTCCAGCGAGGAAGCTATTAATGCTCCGGTATGCC 1260
QY 421 IleIySLeuAlaAspProAlaAspProGluIySgIyLeuMetPheAspGlyArgIleAla 440
DB 1261 ATCAAAATTGGCAGATCTCGCCATCTTCAGAAAGCTGTGATTTGACGCTGAAATGCT 1320
QY 441 GluAspPheIySLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
DB 1321 GAAAGCTTCAAGCTTCTTCCAGGGGATTTGTTCAGGCTTTGGGCAATGGCGAGCGG 1380
QY 461 ValLeuGluGlyIySerTyrValLeuAspValAlaAlaAlaProAspArgGluCyS 480
DB 1381 GTTTCGAAAGGGCGGCTTACAGTCCGACGTAAGTGTTGCTGCTCATCGTGAATGC 1440
QY 481 LeuGlyIySLeuValPheProArgLeuLeuAspCyArgAlaIleuSerGlyLeuGlyIyS 500
DB 1441 CTTGGAATGCTGTGTCTCCGCTCTTTCGACCTGCGCGCTTGTCCGGGCTTGAAGAAA 1500
QY 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTyrPheAlaAspTyr 520
DB 1501 GAGCGCTCGAGCGCGGAGGCTTGGCACTGAGCGCGGCTTGGGCTTGTGCTGACTGG 1560
QY 521 LeuIySArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeu 540
DB 1561 CTCAAAACGACTCAATCGAAGAACCACTGGCAATGCAAGCGCATCATGTGGGTAGGCTTC 1620
QY 541 LeuAspThrProProSerIleAspIySgIyGlyValAlaThrAspIySgIySerIleGln 560
DB 1621 CTCGATACCGCCCGCTGATTAAGGGCGAGGCTACGACAAAGGCTCGATCAACCG 1680
QY 561 ArgAlaValLeuGlnTyrPArgSerAlaIySValAspAlaLeuTyrArgIyGlyIySgIn 580
DB 1681 CGCGCTTTTGCATATGGCGGCTGGGAAAGTTGATCCCTTATCTGTGTGAATCAA 1740
QY 581 SerMetLeuArgAspGluAlaThrLeu 589
DB 1741 TCCATGCTGCGTGAAGGCGCACACTG 1767

```

RESULT 2
PSP238746

LOCUS PSP238746 4166 bp DNA linear BCT 05-NOV-1999
DEFINITION Pseudomonas sp. vdh (partial), fcs, aat and mac (partial) genes,
strain HRI99.
ACCESSION AJ238746
VERSION 1
KEYWORDS aat gene; beta-ketochiolase; fcs gene; feruloyl-CoA synthetase; mac
gene; methyl-accepting chemotaxis protein; vanillin dehydrogenase;
vdh gene.
SOURCE Pseudomonas sp.
ORGANISM Pseudomonas sp.
Bacteria; Proteobacteria.
REFERENCE
1 Overhage, J., Priefert, H. and Steinbuechel, A.
Biochemical and genetic analyses of ferulic acid catabolism in
Pseudomonas sp. Strain HRI99
J. Appl. Environ. Microbiol. 65 (11), 4837-4847 (1999)
JOURNAL MEDLINE 20011220
PUBMED 10543794
2 (bases 1 to 4166)
Priefert, H.
Direct Submission
Submitted (04-MAY-1999) Priefert H., Westfaelische
Wilhelms-Universitaet Muenster, Institut fuer Mikrobiologie,
Corrensstrasse 3, Muenster, D-48149, GERMANY
FEATURES
source
1..4166
/organism="Pseudomonas sp."
/mol_type="genomic DNA"
/strain="HRI99"
/db_xref="taxon:306"
/country="Indonesia"
1..170
/gene="vdh"
<1..128
/gene="vdh"
/function="NAD dependent oxidation of vanillin to vanillic
acid"
/codon_start=3
/evidence=experimental
/transl_table=11
/product="vanillin dehydrogenase"
/protein_id="CAB60225.1"
/db_xref="GI:6273621"
/db_xref="GOA:Q9RLD0"
/transl_table="TrEMBL:Q9RLD0"
/translation="QMPFGVSSVSGSGFSRASISHTQLRWLTITQNGRHYPI"
138..170
/gene="vdh"
/function="putative terminator"
join(373..378,388..2157)
/gene="fcs"
373..378
/gene="fcs"
388..2157
/gene="fcs"
/gene="fcs"
/function="ATP dependent conversion of ferulic acid to
feruloyl-Coenzyme A"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="feruloyl-CoA synthetase"
/protein_id="CAB60226.1"
/db_xref="GI:6273622"
/db_xref="GOA:Q9RLD9"
/db_xref="TrEMBL:Q9RLD9"
/translation="MRSLEALLPFGRIERLEHMAKTRPEQTCVAAARANGEMRIS
VAMFNVAIAQSILPYGLSARPLIYSGNDLEHLOLAFGMYAGIPLYCPSPAYS
LISODIAKLRHIVGLQGLVFAADAAPQRALETILPDVPAIFRGLAGARTVSP
DSLEPGGI BADNAAPATSPDTIAKPLFTSGSTKPKAVPTTQRNLCAQOQLDTF
PVFGEPPVLDVLPNNHTFGSHNIGIVYNGGTYLLDQKFTAGSFATLNLSET
SPPAVLTVPKGMELVGLERDLREPRARKLPFFAAGISQGIWRLDLVAQOH
CGRRIMMAGLGTETAPSTCTTGGPLSMAGYGLPAPGCEVLVVDGKLEGRFPG
HMSGYWRAPEQNAQAFDEGGYCSGDATKLDADPADPQGLMFGRIADBFKLSGVF

VSVGLPRTAVLEGSSVLDVVVAADRECLGLVFPRLDCLRALSGLEKASDAEVL
ASPPVAFADMLKRLNREATGNASIMVGLDTPSPIDKSVTKGSINQRAVLQW
BSAKVDALYRGEDQSLRDEATL"
2143..2146
/gene="aat"
/gene="aat"
2143..2146
/gene="aat"
2154..3449
/gene="aat"
2154..3449
/gene="aat"
/codon_start=1
/transl_table=11
/product="putative beta-ketochiolase"
/protein_id="CAB60227.1"
/db_xref="GI:6273623"
/db_xref="TrEMBL:Q9RLD8"
/translation="MSWSGAVSAFSDTLVAAVRTPTWIDCGALSLVSPRIDGVKA
REVLMRASLEPQWDSVLASMAQAFDLYLLPRHIGLXGVKSVPAIGVRICTG
FILLRQAGROIQGAADHYLCVAESMSRPIASVTRGGPRIGAPVEPDLMEALFD
PARGDMITATERNLARLYGITTREANSYAVSSPERALPAOEKWIQOETVAVTDQED
LGSYNSRAIEPRKAKLLVTVIRGLAVEALSRLKPVHSGVQTIKNSCAVVDGAA
ALVARSSTOPLYLARILATSVGLEPEHMGGLPAPARILRLARSLSURDIDLFEIN
EAQAQVLAVOHELGIHSHKLNIMGAILGHPPLAATGRLCWTTLAHOLOQANNFRYGI
ASACIGGGQGMVLTLENPHFGSSSARSMINRVDPYLS"
3521..4166
/gene="mac"
3521..3525
/gene="mac"
3535..4166
/gene="mac"
/codon_start=1
/transl_table=11
/product="putative methyl-accepting chemotaxis protein"
/protein_id="CAB60228.1"
/db_xref="GI:6273624"
/db_xref="TrEMBL:Q9RLD7"
/translation="MISFAPMAESLGVOKALALAPALVLCVGLIVNGTGFVHTLSG
LYEKSLINSELRAKTOELKYLQORALPIADEGSIKORSLISQVLAENVADILDPFO
RGRSEILKFRASSREASYSIEVGSNAADKLQSGESDLMADKKNLYEEDLSAV
NALMLGLIEDONEKVELIYIMGLGVTTLTMLSAYSVPWFSRQLVPELAKST"

ORIGIN

Alignment Scores:
Pred. No.: 2,86e-210
Score: 3082.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 1
Gaps: 0

US-09-750-986D-30 (1-589) x PSP238746 (1-4166)
QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyArgGlleuGluArgLeuGluHis 20
DB 388 ATCGCTTCTCTCGAGCGCTTCTTCCCTTCCCGGCTCGAATTTTGAAGCTCTCGAGCAT 447
QY 21 TTPAlaLysThrArgProGluGlnThrCysValAlaAlaArgAlaAlaArgAlaGluTTP 40
DB 448 TGGGCTAAGCCCGCTCAGAACAACTCGTGTGCTGCAGCGCGGCAATGGGGAATGG 507
QY 41 ATGArgTLLSerTyrTrpArgLysMetPheHisAsnValArgAlaAlaLeuArgInsSerLeuLeu 60
DB 508 CGTGTATCGACGTACCGGAAAGTTTCCACAAAGCTCCGGCCCATCGCACAGACTTGCTT 567
QY 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuAlaValSerGlyLysAsnProLeuGluHis 80
DB 568 CTTTACGAGCATACCGGACAGCGCTCGCTGCTTATGCTCTCGAATATGACCTCGAATCAT 627
QY 81 LeuGluLeuAlaPheGlyAlaMetTyrAlaGlyTLeProTyrCysProValSerProAla 100
DB 628 CTTACGCTGCAATTTGGGGCTATGATATGGCGGCAATTCCTATTTGCCCGGTCTCTGCT 687

QY 101 TyrSerLeuLeuSerGlnAspLeuAlaValLeuArgHisIleValGlyLeuLeuGlnPro 120
 DB 688 TATTCACTGCTGCGCAAGATTGGCGAAGCTCGTCACATCTGAGTCTTTGCAACCG 747
 QY 121 GlyLeuValPheAlaIleAspAlaIleProPheGlnArgAlaIleGlnThrIleLeuPro 140
 DB 748 GAGCTGCTCTTGCTCCCAATGACGACCTTTCACGCGCAATGACATCTTCCGCG 807
 QY 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPhe 160
 DB 808 GAGGAGTGCCCGCATCTTCACTCGAGCGAATTTGGCCGCGCGCACGCGAGATTTT 867
 QY 161 AspSerLeuLeuGlnGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaIleThrGly 180
 DB 868 GACAGCTCTGAGACGCTGGTGGGATTTGAGCGCATATATCTTTGGCGCACTGGC 927
 QY 181 ProAspThrIleAlaIlePheLeuPheThrSerGlySerThrIleLeuProValAlaVal 200
 DB 928 CCGGATACGATTGCCAAGTCTTGTCATCTTGGCTTACCAACTGCTTAAGCGCGTG 987
 QY 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnGlnMetLeuLeuGlnThrPheProVal 220
 DB 988 CGACCTACTCAGGAGATGCTCTGCGCATACAGCAGATGCTTTCGAACTTTCCGGTT 1047
 QY 221 PheGlyGluGluProProValLeuValAspTTPLeuProTTPAsnHisThrPheGlyGly 240
 DB 1048 TTTGGTGAAGAGCCCGCTGGTGGTGAATGCTGCTTCCGTGGAACCACTTCCGCGC 1107
 QY 241 SerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrIleLeuAspAspGlyVal 260
 DB 1108 AGCCACACATCGCGCATCGTGTGTACACGCGCGCACGTAACCTTGAACGCGGTAA 1167
 QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluLeuSerProThrAla 280
 DB 1168 CCAACCGCCCAAGGGTTCGCGAGACGCTTCCGCACTTGAAGCAATCTCTCCACTGG 1227
 QY 281 TyrLeuThrValProIleGlyTTPGluGluLeuValGlyAlaLeuGluArgAspSerThr 300
 DB 1228 TACCTCAGTGTGCGAAGGCTGGGAGGAATTAAGTGGTGCCTTGAAGCGACAGTACC 1287
 QY 301 LeuArgGluArgPhePheAlaArgMetIleLeuPhePhePheAlaIleGlyLeuSer 320
 DB 1288 CTGGCGCAACGCTTCTTCGCTGCATGAAAGCTGTCTTCTTCGCGCGGTGGGTGTG 1347
 QY 321 GlnGlyIleTTPAspArgLeuAspArgValAlaGluGlnHisCysGlyGluArgIleArg 340
 DB 1348 CAAGGATCTGGGATCTGTTGACCGGGTCCGCGAAGACAGCTGTGTAGCGCATTTCC 1407
 QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360
 DB 1408 ATGATGGCGGGTCTGGGCAATGACGAGACTGCTCTTCTGCACTTTTACACCGGACCG 1467
 QY 361 LeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGluValIleValLeuValPro 380
 DB 1468 CTGTGCATGCTGGTTACATTTGGCTGCGACGCGCTGGCGAGGTCAAGCTCGTTCCG 1527
 QY 381 ValAspGlyLeuLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTTPArg 400
 DB 1528 GTCGATGGGAAATTTGAGAGGGCGTTTCCATGCTCCGACGTCATGAGCGGCTACTGCG 1587
 QY 401 AlaProGlnGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAla 420
 DB 1588 GCTCTCGAACAAATGCCCAAGGCTTCCAGAGGAGGCTATTAATGCTCCGCTGATGTC 1647
 QY 421 IleValLeuAlaAspProAlaAspProGlnIleValLeuMetPheAspGlyArgIleAla 440
 DB 1648 ATCAAAATTGGCAGATCTGCGCATCTCAAGAAAGGCTGATGTTGACGCTCAATTTGCT 1707
 QY 441 GluAspPheIleLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
 DB 1708 GAAAGCTTCAAGGTGCTCAAGGCGTATTTGTACAGGTTGGGCGCATTTGGCGACGCGGG 1767
 QY 461 ValLeuGluGlyGlySerTyrValLeuAspValValAlaIleProAspArgGluCys 480

DB 1768 GTTCTGGAAGGGCGCTTACGTCCTGAGACGTAAGTGTGCTGCTTATCTGTAATGC 1827
 QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyVal 500
 DB 1828 CTTGATGCTGCTGTTTCCGGTCTTCTTCGACTCCGCTGCTTTCGCGGCTAGGAAA 1887
 QY 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaIlePheAlaAspTTP 520
 DB 1888 GAGCGCTCGACCGCGAGGTGCTTGCAGTGAAGCGGTTCCGGCCCTGTTGCTGACTGG 1947
 QY 521 LeuValArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTTPValGlyLeu 540
 DB 1948 CTCMAACGACTCATCTGACAGAACACTGGCAATGCGCATCATGTGGGTAGGGCTTC 2007
 QY 541 LeuAspThrProProSerIleAspIleValIleValThrAspIleGlySerIleAsnGln 560
 DB 2008 CTGATACCGCCCGCTGATTTGATTAAGGCGAGGCTACTGACAGGCGCTGATCAACCG 2067
 QY 561 ArgAlaValLeuGlnTTPArgSerAlaValAspAlaLeuTyrArgGlyGluAspGln 580
 DB 2068 CCGGCTGTTTTCGAATGCGGTCGCGGAAAGTTGATGCCCTGTATGTGTGTAAGTCA 2127
 QY 581 SerMetLeuArgAspGluAlaThrLeu 589
 DB 2128 TCCATGCTGCGTACGAGGCCACACTG 2154
 RESULT 3
 A92090 LOCUS A92090 32679 bp DNA linear PAT 22-JAN-2000
 DEFINITION Sequence 1 from Patent EP0845532.
 ACCESSION A92090
 VERSION A92090.1 GI:6740886
 KEYWORDS
 SOURCE Pseudomonas sp.
 ORGANISM Pseudomonas sp.
 Bacteria; Proteobacteria.
 REFERENCE 1 (bases 1 to 32679)
 Priester, H.D. and Rabenhorec, J.D.
 Enzymes for the synthesis of conferyl alcohol, conferyl aldehyde,
 fennilic acid, vanillin, vanillic acid and their applications
 Patent: EP 0845532-A 1 03-JUN-1998;
 JOURNAL HAAHMAN & REIMER GMBH (DE)
 FEATURES
 source Location/Qualifiers
 1..32679
 /organism="Pseudomonas sp."
 /mol_type="unassigned DNA"
 /strain="HR199"
 /db_xref="taxon:306"
 gene 3146..4000
 /gene="ORF1"
 CDS 3146..4000
 /gene="ORF1"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB69475.1"
 /db_xref="GI:6740887"
 /translation="MIAITGASGQGRITLALRLRPASEIIALVRDPAKADLLAR
 GIVRQADYRREPTELALIGVNRLLIISSEVGORTQHRVITAQKQEGELAYT
 SLVHDKSLGATSHRDTQALTESGIPHYLLRGMWTHENTAGI PVALVGVLLC
 KODGILASARADYRPAALAVITGENOKRVELNGEPAYTTTELAAYAPQAGTIV
 YSNLSQDYRSRLISAGLPDGPALLADSDAAGAGYLFDSGSDSRKLI GRPTTWS
 AIAAAIGR"
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,2e-209 Length: 32679
 Score: 3082.00 Matches: 589
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

us-09-750-986d-30 (1-589) x A92090 (1-32679)

QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyValArg11LeuGluAlaLeuGluGluHis 20
 Db ATGCGTCTCTCCAGAGCGCTTCTTCCCTCCGCGGCTGAATCTTGAGCCCTCCGAGCAT 21347
 QY 21 TrpAlaLeuThrArgProGluGluGlnTrpCysValAlaAlaArgAlaAlaAsnGlyGluTrp 40
 Db TGGGCTTAAGACCCGCTCCAGAACAACTTGCTGTCTCCAGGGCGGCAATGGGAAATGG 21407
 QY 41 ArgArg15SerTrpAlaGluMetPheHisAsnValArgAla11LeuGlnSerLeuLeu 60
 Db CGTGGATACGATACGCGGAAATGTTCCACACGTCGCGGCATCGCAGAGAGCTTGCTT 21467
 QY 61 ProTrpGlyLeuSerAlaGluArgProLeuLeu11ValSerGlyAsnAspLeuGluHis 80
 Db CCTTACGGAATACGAGAGAGCGCTCCCTTATCTCTCTCGAAATGACCTGGAAACAT 21527
 QY 81 LeuGlnLeuAlaPheGlyAlaMetTrpAlaGly11LeuProTrpCysProValSerProAla 100
 Db CTTCACCTGGCATTGGGCTTAATGATGCGGCACTTCCATATGCGCGTGTCTCTGCT 21587
 QY 101 TrpSerLeuLeuSerGlnAspLeuAlaValLeuArgHis11ValGlyLeuGlnPro 120
 Db TATTCACTGCTGCGCAAGATTGGCGAGACTGGTCAATCGAGGCTTCTGGCAACCG 21647
 QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAla11LeuTrp11LeuPro 140
 Db GGACTGCTCTTCTCCGATGACGACCTTTCAGCGCCCAATGAGACCATCTGCGC 21707
 QY 141 AspAspValProAla11PheThrArgGlyGluLeuAlaGlyArgArgTrpValSerPhe 160
 Db GAGGAGCGTCCGCAATCTTCACTCGAGGGAATGGCGGCGGCGCACGGAGATT 21767
 QY 161 AspSerLeuLeuGluGlnProGlyGly11LeuAlaAspAsnAlaPheAlaAlaThrGly 180
 Db GACAGCTGCTGAGCAGCTTGCGGGAATTGAGGCAATATGCTTGGCGCACTGGC 21827
 QY 181 ProAspThr11LeuAlaValPheLeuPheThrSerGlySerThr11ValLeuProValAla 200
 Db CCGAATACGATTCGCAAGTCTTGTTCACCTTGGCTTACCAACGCTTAAGCGCGT 21887
 QY 201 ProThrThnGlnArgMetLeuCysAlaAsnGlnGlnMetLeuGlnTrpPheProVal 220
 Db CCGACTACTACGGAATGCTCTGGCCATCAGCAAGATGCTTCTGCAACCTTCCGCT 21947
 QY 221 PheGlyGluGluProProVal11LeuValAspTrpLeuProTrpAsnHisThrPheGly 240
 Db TTTCGTGAAGAGCGCGCGTGTGGTGAATGGTGGCTGGGAAACCACTTCGCGCGC 22007
 QY 241 SerHisAsn11LeuGly11ValLeuTrpAsnGlyGlyThrTrpLeuAspAspGlyVal 260
 Db AGCACAACATTCGCGATCGTGTGTACACGCGCGCACGTAATCACTTTCGACGCGTAAA 22067
 QY 261 ProThrAlaGlnGlyPheAlaGluTrpLeuArgAsnLeuSerGlu11SerProThrAla 280
 Db CCAACGCCCAAGAGGTTCCGCGAGACGCTTGGCACTTGCGAAATCTTCCACTGCG 22127
 QY 281 TrpLeuThrValProValGlyTrpGluGluValGlyAlaLeuGluArgAspSerThr 300
 Db TACCTACGTGTGCGAAAGCTGGAGGATTAATGTGGTCCCTTGAAGCAAGTAC 22187
 QY 301 LeuArgGluArgPhePheAlaArgMetValLeuPhePhePheAlaAlaAlaGlyLeuSer 320
 Db CTCGCGGACGCTTCTTCCCTCCATGAACTGTCTTCTTCCGCGCGCGTGGGTTGTCG 22247
 QY 321 GlnGly11LeuTrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGluArg 340
 Db CAAGGAGCTTGGGATCGTTTGACCGGGTGTGTAAGCACTGTGTGAGCGCATTCGCG 22307
 QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrTrpGlyPro 360
 Db ATGATGGCGGCTGTGGGCAATGAGAGACTGCTCTTCTTTCACCGAGCGG 22367

QY 361 LeuSerMetAlaGlyTrp11LeuGlyLeuProAlaProGlyCysGluVal11LeuValPro 380
 Db CTGTCAATGGCTGTATACATTGGGCTGCGACGCCCTGGCTGGCAAGCTTCGCTCCG 22427
 QY 381 ValAspGlyValLeuGluGluGlyArgPheHisGlyProHis11ValMetSerGlyTrpTrpArg 400
 Db GTTCATGTGGAAATTTGGAAGGCGCTTTCATGTGCTCCGACCTCATGAGCGGCTAC 22487
 QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTrpTrpCysSerGlyAspAla 420
 Db GCTCTGAACAAATATCCCAAGCGTTTCSACGAGAAAGGCTATTAATGCTCCGCTGATGCC 22547
 QY 421 11LeuLeuAlaAspProAlaAspProGln11ValGlyLeuMetPheAspGlyArg11LeuAla 440
 Db ATCAAAATGGCAATCCCTGCGATCCCTCGAAGAAAGTGTGATGTTGAGGTCGAATTCCT 22607
 QY 441 GluAspPheValLeuSerSerGlyValPheValSerValGlyProLeuArgTrpArgAla 460
 Db GAAAGCTTCAAGCTGTCTTCAAGGGGATTTGTCAAGCGTTGGCCATTGGCGACGCGCGC 22667
 QY 461 ValLeuGluGlyGlySerTrpValLeuAspValValAlaAlaAspAspArgGluCys 480
 Db GTTCTGAAGCGCGCTCTTACGTCCTGACGTAGTGTGCTGCTCTCTGATGATGC 22727
 QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyVal 500
 Db CTTCGATGCTCGTGTTCGCGCTTCTGACCTGCGCTTGTGCGGGCTGAGGAAA 22787
 QY 501 GluAlaSerAspAlaGluVal11LeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
 Db GAGCGCTCGAGCCCGAGAGCTTTCGCAAGACCGGCTGGCGCTGCTGACTCG 22847
 QY 521 LeuValArgLeuAsnArgGluAlaThrGlyAsnAlaSerArg11MetTrpValGlyLeu 540
 Db CTCAAAAGCACTCAATGAGAGCAACTGGCAATGCCAGTGCATCATGTGGTGAAGGCTC 22907
 QY 541 LeuAspThrProProSer11LeuAlaValGlyVal11ThrAspValGlySer11LeuGln 560
 Db CTGATATCGCGCGCTGATGATTAAGGGGAGGCTCACTGACAAAGGCTGATCAACAG 22967
 QY 561 ArgAlaValLeuGlnTrpArgSerAlaValAspAlaLeuTrpArgGlyGluAspGln 580
 Db CGCGCTGTTTGAATGCGCGTGGCGGAAGTTGATCCGTGATGTGTGAAGATCAA 23027
 QY 581 SerMetLeuArgAspGluAlaThrLeu 589
 Db TCCATGCTGCGTGAAGGACCACTG 23054
 QY 23028 TCCATGCTGCGTGAAGGACCACTG 23054

RESULT 4
 AEO16866/c
 LOCUS 310029 bp DNA linear BCT 09-OCT-2003
 DEFINITION Pseudomonas syringae pv. tomato str. DC3000 section 11 of 21 of the complete genome.
 ACCESSION AEO16866 AEO16853
 VERSION AEO16866.1 GI:28853240
 KEYWORDS
 SOURCE
 ORGANISM
 Pseudomonas syringae pv. tomato str. DC3000
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
 1 (bases 1 to 310029)
 Buell, C.R., Joardar, V., Lindeberg, M., Selengut, D., Paulsen, I.T., Guim, M.L., Dodson, R.J., Deboy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Daugherty, S., Brinkac, L., Beaman, M.J., Haft, D.H., Nelson, W.C., DavidSEN, T., Zafar, N., Zhou, L., Liu, J., Yuan, Q., Khouli, H., Fedorova, N., Tran, B., Russell, D., Berry, K., Khouri, T., Van Aken, S.E., Feldblyum, T.V., D'Ascenzo, M., Deng, W.L., Ramos, A.R., Alfaro, U.R., Cartimoun, S., Chatterjee, A.K., Delaney, T.P., Lazarewitz, S.G., Martin, G.B., Schneider, D.J., Tang, X., Bender, C.L., White, O., Fraser, C.M., and Collier, A.
 The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000

JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 100 (18), 10181-10186 (2003)
PUBMED	12928499
REFERENCE	2 (bases 1 to 310029)
AUTHORS	Berry, R., Jorjadar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K., Uterback, T., Van Aken, S., Feldblum, T., Gwin, M., Dodson, R., DeBoy, R., Durkin, A., Kolonay, J., Madupu, R., Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J., Nelson, W., Davidsen, T., White, O., Fraser, C. and Collier, A.
TITLE	Submitted (03-MAR-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1. .310029
gene	/organism="Pseudomonas syringae pv. tomato str. DC3000"
gene	/mol_type="genomic DNA"
gene	/strain="DC3000"
gene	/db_xref="taxon:223283"
gene	/note="pathovar: tomato"
gene	204. .1664
gene	/locus_tag="PSPTO2809"
gene	204. .1664
gene	/locus_tag="PSPTO2809"
gene	/note="similar to GP:15140142; identified by sequence similarity; putative"
gene	/codon_start=1
gene	/transl_table=11
gene	/product="conserved hypothetical protein"
gene	/protein_id="AA056309.1"
gene	/db_xref="GI:28853241"
gene	/translation="MSDVAAPAGLAAALAEARLQODLAWLDPAAKPKPPNAGAVTDVAIGGMAGLALAAELRHGAAVAFDOSPAPFEPMPATTAMETLRSEKQITGALGAPALTFPAMTEAQFGADGWAALDKIPQWMAYLKWRKALDVRNEHRSRVAPRTDGLVLDITMPGQTRFMLARHVATLGRGLGPPWPDFARQLPHLMTWSAAGD DMEGRRVAVTGGGASANDSAATALAGAQVLLIRALPVPNGKGGNGMGVGY YMRPIMMKRIIRHYLNROQVPPRPGSTLRVSSGRARFLNDPVLAVENAPAGVWL QNSKARVADPVRVATGFRTPDELRPEPAPPSHTRVMODRPEAPPEGADSEIAPD LKCFRFOEKTTPACPGNLNHFCSIPALISTGAVSGDIPATSEGSKRLLAHLTGLFP NBDIALHFSFMRVYAPPELLGDWVASEPGETELRP"
gene	1661. .2635
gene	/locus_tag="PSPTO2810"
gene	1661. .2635
gene	/locus_tag="PSPTO2810"
gene	/codon_start=1
gene	/transl_table=11
gene	/product="peptide ABC transporter, permease protein"
gene	/protein_id="AA056310.1"
gene	/db_xref="GI:28853242"
gene	/translation="MTGMLRRLTOSLIVITMTVVVFGVNAIGNPMIIVGEDINO ARLQAIADLGDQPLMOQYLITLKGAMNGNIGOSVYTHEDANRLILQRLPATFELAF SAEFLAVVIGVPLGPMGYPDHPLSRLLMAASIVGFSLPAPVVALMMIMLPSITLGGW LPSRGGETREWLGVQSWMLTDGLQHLPLPALMLALFKISLVLRLTRAGVRSVLEQF EYFARAKGLSPWRVCMHVMRTWMPVTVLAMELGSTIAYAVTESIFAMPGAGKL IDSINMLDRPVPVAVLAVVIVFVVLNIVDGLVYLLDPRVIEASR"
gene	2632. .3552
gene	/locus_tag="PSPTO2811"
gene	2632. .3552
gene	/locus_tag="PSPTO2811"
gene	/codon_start=1
gene	/transl_table=11
gene	/product="peptide ABC transporter, permease protein"
gene	/protein_id="AA056311.1"
gene	/db_xref="GI:28853243"
gene	/translation="MNVIKAVAPANTLLQAQSPWRVIVDFMSSSTAVVGLLLVIV LLAAPAVIVQPYDLMQNLNMDVAMPGLSMDSGTYWLTGDSQGLVSAIYGL RSLWVIGSALIAIVIGTGLVSAVAGWMDALMLRLDILLSPVITAMMLTA WKGKGVNMVLTILLEMAVYVARTAGCATSRARVYDARQGVGPRIYIGHILP NCPLPIVIGALQIARAITLEATLTGAGVPTESPSTGLLIANGQYMLSNBYWISL FEGALITIVAINIVGRLRDLNRLQR"
gene	3549. .4562
gene	/locus_tag="PSPTO2812"
gene	3549. .4562
gene	/locus_tag="PSPTO2812"
gene	/codon_start=1

gene	/transl_table=11
gene	/product="peptide ABC transporter, ATP-binding protein"
gene	/protein_id="AA056312.1"
gene	/db_xref="GI:28853244"
gene	/translation="MSLPVSRVTSVEPTELVRLCTSPYRAGVLPAVRNYSLVQ PGRIQLVSSSGSKSVTFSTLGVYDERGISGGCVLPKRGDITLSTGQAREIGN RAAVFPQDMWTLNPLVRLDTOMTEAVRSHSTLGRKEAEHASTLALWGLSPEERL RYMPHQLSGMQRVAIAIALHAPDLIADEPTTALDVTIAQIISVQKVRQEGT SLIWIHDSVVAGLADVDVAVVAGRIIVEQSVADVLDVDRPHQYTGGLIDSPSRMR GGRLRQIPGMADLLSMPAGCAFAARCAKAPACQDPPREILPGHTVRCFYGAGA DIO"
gene	4552. .5565
gene	/locus_tag="PSPTO2813"
gene	/note="This region contains an authentic point mutation, causing a premature stop, and is not the result of a sequencing artifact; peptide ABC transporter, ATP-binding protein, authentic point mutation; similar to GP:15158466; identified by sequence similarity; putative"
gene	5618. .7183
gene	/locus_tag="PSPTO2814"
gene	5618. .7183
gene	/locus_tag="PSPTO2814"
gene	/note="similar to GP:15140494, and SP:P23847; identified by sequence similarity; putative"
gene	/codon_start=1
gene	/transl_table=11
gene	/product="peptide ABC transporter, periplasmic peptide-binding protein"
gene	/protein_id="AA056313.1"
gene	/db_xref="GI:28853245"
gene	/translation="MKPLVRSLLAGALVLAQSVASQDIRIGYADPVSSLDPOLNYYA GRSVALHFEZLVSRDRTKLPGLAKSKVITDDTWEALREDEVQMGTPLTADL SVFEBARVPSVSAVAGAMRTVSKAKDHTLILIKRLPNANLIPDVSIVYIVR HGAASASADVNSGALICTGPIYRVSFPQGRITFARNDSYGAKPTMDKDFERIA TDMNKPALADVDVIVKASPLSLAIPPAIDERMQSTVNAOMMPANRPGVPEF IKNIPIVKAQDLAQGFPEGQITVHPGDRIPQAEVQAQVAFTRIGYKQVL EYLPAVAVVAGKANFELASVLAWGNGTEAAVALTNLITTVDSKGSQSWGHYSN PLVDKALADSTAEFDEAKRRKILBDSVQVSDVGIILPFIYQNIWDARKGLKVEPLV SDRTAIVWTPEQP"
gene	7255. .7776
gene	/locus_tag="PSPTO2815"
gene	7255. .7776
gene	/locus_tag="PSPTO2815"
gene	/note="similar to GP:15162222; identified by sequence similarity; putative"
gene	/codon_start=1
gene	/transl_table=11
gene	/product="conserved hypothetical protein"
gene	/protein_id="AA056314.1"
gene	/db_xref="GI:28853246"
gene	/translation="MTTESNSPPDLLHLGIEPGNPLHTVHARDKVASATQSGAL FDPDALDNLPLSERLMAVYVATQISAOPTLRAVYIQOLREGVADSVADVANDID LLEDPRIALILPRTILSPHSGQSALQALQGLSTAEIVLAQLIAFLSYQVRL AAGLTAIQSAGAA"
gene	7776. .8366
gene	/locus_tag="PSPTO2816"
gene	7776. .8366
gene	/locus_tag="PSPTO2816"
gene	/note="similar to GP:15162222; identified by sequence similarity; putative"
gene	/codon_start=1
gene	/transl_table=11
gene	/product="alkylhydroperoxidase AhpD domain protein"
gene	/protein_id="AA056315.1"
gene	/db_xref="GI:28853247"
gene	/translation="MSLEIRGFTNQLGWKAMLPYDNLNATLAEQTVAEESHPQ RTSDDYTLAHPDLIRQSRQAFNAIMYAPGSLAEELASTVSRINRCVYCASVH AORPEQLAKRNDVIAQVADPATAGTTRAKELIVQPAIDLLEPASLNAHIIALREQ GLDDAOVLDLHAIATPAMNMLNLTGSPVPEAE"
gene	8429. .8614
gene	/locus_tag="PSPTO2817"
gene	8429. .8614

/locus_tag="PSP02817"
/note="similar to GP:15158870, and GP:15980012; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AA056316.1"
/db_xref="GI:28853248"
/translation="MAECHPSLPMVCARVSGATVTEPRKARTMHTPLQGLIVTNG AGVSDAETASGAPQ"
complement (8591..9982)
/locus_tag="PSP02818"
complement (8591..9982)
/locus_tag="PSP02818"
/note="identified by match to Pfam protein family HMM PF01569"
/codon_start=1
/transl_table=11
/product="fatty acid desaturase"
/protein_id="AA056317.1"
/db_xref="GI:28853249"
/translation="MYLVLTGLVPLVTHVPMGSVRLVQPSGVDAHVPLLPFLPLYL SYLVNPLVVMGRQSWLPTFPAGALAGLCVCHLEFPTMLRBERGAAMLDILY QLDAPLAPSGHVALPVALSVAMAGRLQKAVFWMSVLMLTMTTGOHFTDMA YGAALGVACGAATLIPKRCVLDRLTSALLEMLCLITRVLVLAIDMRFYLLMLI VVAROALFVLVYDATHNVLTERRSTNDPLINLAIGPGLVPIEFYRPLLDHNOAG TEODPERRFLYTRQPRFQPLSKLRLQLGLDLMVNTLRNIAATKAGAPKPTR PLISALVMLMTVACLWQCSIREVGLMALMFLPVVTGLTQKRSIAESHGSGFV TRQMDQMTYAMRVGMVIGRFPIWPIHNLQHRHASIPWHALPSAVGKDEKLMPRSA LPLMWSGLRKR"
10166..10954
/locus_tag="PSP02819"

gene

Alignment Scores:
Pred. No.: 2, 09e-164 Length: 310029
Score: 2474.00 Matches: 460
Percent Similarity: 87.48% Conservative: 50
Best Local Similarity: 78.90% Mismatches: 73
Query Match: 80.27% Indels: 0
DB: 1 Gaps: 0

US-09-750-986D-30 (1-589) x AE016866 (1-310029)

QY 1 MetArgSerLeuGluAlaLeuLeuProRheProGlyValGllLeuGluArgLeuGluHis 20
DB 172246 ATGCGAGCCCTGGACCGGTGGCGCGCTGCGCGATGCGTTCGACCGCGCTGGTGCAC 172187
QY 21 TrpAlaLeuThrArgProGluGluGlnThrCysValAlaAlaArgAlaAlaLeuGluTyr 40
DB 172186 TGGCCAGCGCTGGTGGCGCAACAACCTTTATGTGCCCGCTGACACCGCTGGTGGCTGG 172127
QY 41 ArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
DB 172126 CGCAGAGTCAGTTACCGCGCAAGCTCTACACCGATTCGCCCATTTGCCAGACCTGCTG 172067
QY 61 ProTyrGlyLeuSerIleGluArgProLeuLeuIleAlaSerGlyAsnAspLeuGluHis 80
DB 172066 GCTTACGGGCTTTCGGCAGGCGCGCACTGGCGCTGTGCGGTAAACAATCGACGACAT 172007
QY 81 LeuGlnLeuAlaPheGluAlaMetTyrAlaGlyIleProTyrCysProValSerProAla 100
DB 172006 TTGCAATATCCTTTGGGCGCTATGTACGCGGTATTCCTTATTTGCCGGTGTACCGGCT 171947
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaAlaLeuArgHisIleValGlyLeuLeuGlnPro 120
DB 171946 TACTCGGTCTGTGCGACGACTTCGCGCAAAATTCGGGATGTCTGTGAGGTTCACAGCCA 171887
QY 121 GlyLeuValAlaPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
DB 171886 GGGCTGTGTCTTTCACGAGATGCGCGCGCTTCACGCGCCACATCGATGCTCATACCG 171827
QY 141 AspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyArgArgThrValSerPhe 160
DB 171827

DB 171826 GCTGACACACCGGTGATACGGTATACGGGACAGTTGAATGCGCGGCTCATTCGACCTT 171767
QY 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
DB 171766 GCCAGCTGCTTAAACAACCGGGCGGACGAGGCGGACGCTGTCATGCGCAGCGGG 171707
QY 181 ProAspThrIleAlaIysPheLeuPheThrSerGlySerThrIysLeuProIlyAlaVal 200
DB 171706 CGGACAGCATTCGCCAAGTTCTTTTCACCTCGGGCTCGACCAAACTGCCCAAGACATG 171647
QY 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnIleMetLeuGlnThrPheProVal 220
DB 171646 GTTACCACTACGCGGATGCTGTGCAATCAACAGATGCTGTGCAAACTCCGGGTG 171587
QY 221 PheGlyGlyGluProProValLeuValAspThrLeuProThrAsnHisIleThrPheGly 240
DB 171586 TTGCGGGAACGCGCGCGGTGTGAGATGAGTCCCGTGAATACATCTTCGCGGCG 171527
QY 241 SerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspArgIlyLys 260
DB 171526 AGTCATTAACCTCGGTATCGTCTCTCAACGCGGCGACGCTTCTATCTCGACAAATGCGAA 171467
QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
DB 171466 CCGACTGTTACAGGCTTTGGCGAAACCTTGCGCACTCAAGAAATCTCGCGACAGCC 171407
QY 281 TyrLeuThrValProIlySerGlyTyrGluGluLeuValGlyAlaLeuGluAlaAspSerThr 300
DB 171406 TACCTACCGCTGCGCAAGCGCTGGGAAGCTGTGTAACCGCGTGACGAGCGCGAG 171347
QY 301 LeuArgGlyArgPhePheAlaArgMetIlyLeuPhePhePheAlaAlaIleGlyLeuSer 320
DB 171346 TTGCGTAGCGATCTTCTTCGCGCATGACGCTTCTTCTCCCGCGCGGTTTCTCG 171287
QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGlnGlnHisCysGlyGluArgIleArg 340
DB 171286 CAATGATATCGGAGCCGCTCGACCGGATGGCGGACAGCAGCATGTGGCGGACCATTCGC 171227
QY 341 MetMetAlaGlyLeuGlnIlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360
DB 171226 ATGATGCGCGGTGGGACAGACCGGACCGCGACCGCTTCGACCTTTACACCGGGCGG 171167
QY 361 LeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGluValIlyLeuValPro 380
DB 171166 CTGTCAATGCGCGGTTATCATCGGCTTGGCGCGCGCGGAGTGTGTCTGTGTCGG 171107
QY 381 ValAspGlyLeuLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTrpArg 400
DB 171106 GTTGATGGCAAGCTGGAAAGGCGCTTTCGCGCGCGCACATCATGCTGGCTACTGGCG 171047
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGlnGlyTyrTyrCysSerGlyAspAla 420
DB 171046 TCGCCGACGAAACCGCTGAGGTTCGACGAAACGCGCTTTTATCGTTCCGGTGAACGG 170987
QY 421 IleLeuLeuAlaAspProAlaAspProGlnIlyGlyLeuMetPheAspGlyArgIleAla 440
DB 170986 ATCAAGCTGGACAGACCTTTCGCGCGCGGACCTGAGCTATGTTCATGAGGCTTGGCC 170927
QY 441 GluAspPheIlyLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
DB 170926 GAGGATTTCAAACTGTCTCGGCTGTGTTGTCTGACGCTGGGCAATCGCAATCGGCG 170867
QY 461 ValLeuGluGlyIlySerTyrValAlaAspValAlaAlaAlaProAspArgGlyCys 480
DB 170866 GTCTCGAAGGCTCGCGGATGTTCAGAGATGTGTATCACCGCACCGGATTCGAGTGC 170807
QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaIleLeuSerGlyLeuGlyLys 500
DB 170806 CTGGGGCGGCTGTCTTCTCTCCCTGCGACAGATCCGACAAATTCGCTGGCTGGGTGCG 170747
QY 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
DB 170746 CAGCGCAGCATGTCAAGATTTGGCCAGACCGCGGTACGCGAGTGTTCGCGACTGG 170687

QY	521	Leu1ySaAqgleuAsnAArgClua1aThrG1yAenA1aSeAArg1lEmeTtrPValG1yLeu	540
Db	170686	CTGGACACCTTGAATGAAGAGGGCCAAATGGCAACGCCAGTCGCTTGAAGTGAATCGGGCTG	170627
QY	541	LeuAAPTTrProProSerIleAsp1ySg1yG1uVal1ThAAsp1ySg1ySer1leAenG1n	560
Db	170626	CTGATGTAACCAAGCCTCAATTCATCGCGCCGAATACTGACAAAGGCTCTATCAACCAAG	170567
QY	561	Arg1aVal1leuG1nTTrPArgSer1a1ySVal1aPa1leuT1yAArg1yG1uAePg1n	580
Db	170566	CGTGGCGGTGTGCACTGGCGGCCAGAAAGTCGAAGCCGTGATGCATCAGATGCGC	170507
QY	581	SerMet1eU 583	
Db	170506	TCACGGCTC 170498	
RESULT 5			
LOCUS	PFL536325	9477 bp	DNA linear BCT 09-JAN-2004
DEFINITION	Pseudomonas fluorescens feRR gene, ech gene, vdh gene, fcs gene, aat gene and fcd gene.		
ACCESSION	AJ536325		
VERSION	AJ536325.1	GI:40788116	
KEYWORDS	aat gene; beta-ketothiolase; ech gene; fcd gene; fcs gene; feRR gene; fer1oy1-CoA dehydrogenase; fer1oy1-CoA hydratase; fer1oy1-CoA synthetase; transcriptional regulator; vanillin dehydrogenase; vdh gene.		
SOURCE	Pseudomonas fluorescens		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.		
REFERENCE	1		
AUTHORS	Barghini, P., Civalami, C., Picea, A. G., Schiesser, A. and Ruzzi, M.		
TITLE	Cloning and Characterization of Ferulate Catabolism Genes from Pseudomonas fluorescens Bp13		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 9477)		
AUTHORS	Ruzzi, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-JAN-2003) Ruzzi M., Agrobiology and Agrochemistry, DABAC - University of Tuscia, via C. de Vellis, snc, Viterbo - 01100, ITALY		
FEATURES	Location/Qualifiers		
SOURCE	1..9477		
	/organism="Pseudomonas fluorescens"		
	/mol_type="genomic DNA"		
	/strain="Bp13"		
	/isolation_source="soil"		
	/db_xref="taxon:294"		
	/environmental_sample		
	/country="Italy:Lazio"		
	/complement(1297..1785)		
	/gene="feRR"		
	/complement(1297..1785)		
	/gene="feRR"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		
	/product="transcriptional regulator"		
	/protein_id="CAD60265.1"		
	/db_xref="GI:40788117"		
	/translation="MAKPAPLADPRAPANAEVAPLADSLDELIGAMRPAOLKLF		
	ONLIGQSLAHDIRPAQPSALATIDSNPGMQLDRLRALAIIEPPQVPIPLNKLESRLA		
	QDVRKPKRSTYGIPLSTGTLLKELKQIAQSDVDATSLSAEREBLLRLKKVY		
	1994..2824		
	/gene="ech"		
	1994..2824		
	/gene="ech"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		

gene
CDS
2921. .4369
/gene="vch"
2921. .4369
/gene="vch"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="vanillin dehydrogenase"
/protein_id="CAD60267.1"
/db_xref="GI:40788119"
/translation="MSNYEGRWTVKVEIDEGIAWVILNPEKKNMSPTLNREIMIDV
LETLDEDPAGVVLVTGAEAWTAGMDKEYFREVDAPRELIOEKIREASQOMQKLL
RWYAPPTIIMWCMFCGGSPSLVACDAICADEATFSEIINGNIPGNIVASAMAD
TVGHOSLITVIMGTGTPGGQKAEMGLVNDSPVLAIREVITELARNLEKNPVLR
AAHGKRCRCELWEQNEEDYIVAKNDQSLDLTDEGRBEQMKQPLDDKSIRKGLQAYKK
"
2921. .4369
/gene="vch"
2921. .4369
/gene="vch"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="vanillin dehydrogenase"
/protein_id="CAD60267.1"
/db_xref="GI:40788119"
/translation="MLDVPILLIGQSGRPARDGRTLERNPVTGAVSVSAATLDDDD
AAVAAQAFPMALAPNERSRLKKAEOLOARSGEFIAAAGETGMANWGENPVR
ILAAMLREASSTWTOINGEVISDIPGSPFAMALQPGCVGIVAPMAVPLALRAIT
MPLACNTPYVLKASLSLSPAVHRLTGOVLDGQDGVVYVSNAPVDAALVEVLIAN
PVARRNPTFGSTHVGRIVGELEAARHLKPLLLELQKAPLVLDLDDDLDAVAAALFG
YNNQOQICMSTERLIVDAKVADAPAKAAVATLRADPBAASVLSGLVDSAGRG
IKALIDDLASIKARLVVGQLDGSLLOPTLLDGVTEWRILYRESFGVAVLIDSGGR
BELLRANDSEFGLSALIFSRDTGALALQARVESGICHINGPTVHDEAQMPPGVKXS
SEYSGFGKASITHEFTQLRWVTLQHGPRHYI"
4628. .6397
/gene="fcs"
4628. .6397
/gene="fcs"
/codon_start=1
/transl_table=11
/product="feruloyl-CoA synthetase"
/protein_id="CAD60268.1"
/db_xref="GI:40788120"
/translation="MRSUKPLAALPDRLLDRLVHMARVREPQGTIAREAGCDNRILVS
YCOADSVAKINOGILRYGLVSDKPLALLSNDIEHLALGAWYAGIPYCPVPAVS
ILSOPFAIRHYCDLLQPLGVSPDASAYQRIADVLPETPLISVRQVGRQAS
ALLBPGGAEDDAFAATGPDPSIAKPLFTSGSTPLPKAVITTOIMLCANQMLKET
PVGPEGPVLYNDMLPMNHTPGSSHVYGLVYNGGTPYLDGKPTAQGPAETLRNLKE
STFATVLYPKKHELVNMLEDDALEBCEFFPKMSLFFFAAGLSQSVWDRLDVLEOH
CEERIIMWAGLMTFAAPSCTFTTGSPJASGLIIPAGCEVRLPVDPGKLEGFREG
HIMPQYVRAAQQTAEVFAQGYFSCGDLAKIADAPDLQGLMFGRIAEDEKLSGVH
VSYGLPRLNAYLEGAPEVQODLVVAAPDECEGLAVFPRLYECRRLAIGASASEAVY
AAPVPRQMGWILCINREASGNARLEWIMALQVEPASIDRGEITDKSINGRAVLQW
RAEOVLYNRYGEPSEIILAEPRK"
6394. .7641
/gene="aat"
6394. .7641
/gene="aat"
/codon_start=1
/transl_table=11
/product="putative beta-ketocholalase"
/protein_id="CAD60269.1"
/db_xref="GI:40788121"
/translation="MSGAGQYSGEADVAIIEAVRTPWVDLGGDLAETSPDLGIKVGGR
EVLAAEPQAVPQVDSVLAGSMAQASFDAYLLPRHIGLISGVQAVPALGAVORTATGEG
ELLRLAABOELRGEVOLALCVASBSWNRPIIAYTRIGCGFRIGAAVQDFIMELVYDPE
ALGLMIAITADNLARLYGTREAVDVGYSCHORLALARGOMFPAETIVASQORFEL
DQYQARGRLTPRGVERVSODSHPRSESDIARLRTIHGGVOTRAGNSCAVDDAAAAA
LYGRASGSDRRLALPASAYVGVAPRWEMQIGRAPATILLORCKPAPAGFOIRSAFPI
NNAQASQVLAVALFELGDGDRILNVQGSGLAHGPLMASGLVLTLLARQLREHNRVIG
IAACVGGQGGQALLIENPAMRA"
7701. .9455
/gene="fcd"
7701. .9455
/gene="fcd"
/codon_start=1
/transl_table=11
/product="putative feruloyl-CoA dehydrogenase"
/protein_id="CAD60270.1"


```

/db_xref="GI:40788122"
/translation="MMSYQAPLADMOFVLEHMLAPBAMRSPAFALDPLALQYLE
OARPSQVILAPLNASGDRCRMADQVSTPSPGADATRAYEGECPALACBAALG
OGIPQLDLALQEMLYASNAMMYTQVLAIGATLYCKTHAPMLQERYPALISGLSL
PMCLTEPLSLASVGLRLCAEPQADQSYLSKLFISGEHDLTDNLHLVLAIP
DAPEGSGSLFLVPRLEQDNGVRCNGIEHKMGFKSATGSLYFDAQGLYVBA
NGLAMFVMMNSARLHVLQGLGVEAAMQNALVAAGEOQKARPLRPAQAQAD
PIRYHAPMRVULLELRATSEGLAIGVMAHLDAOEHNEPEPLARQADLQDLPI
ITAFTEQGRQASNLQVTFGSGYVAERPIEDTLDSRIAMYESNEIQANDLLR
KVLGDGQGFALBQLRDBAQALVAGCAGRPADPLAEALQGLVGEVQWALBD
AEYPRADGDFLALCGLSLALAMARAARVSRNLPGBDPLRGAKLTAFGFFSYLLPQ
VEQRMTAVSAQALAFI"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1,35e-164 Length: 9477
Score: 2449.00 Matches: 456
Percent Similarity: 87.03% Conservative: 54
Best Local Similarity: 77.82% Mismatches: 76
Query Match: 79.46% Indels: 0
DB: 1 Gaps: 0

```

US-09-750-986d-30 (1-589) x PFLS36325 (1-9477)

```

Oy 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyYArgIleLeuGluAUAArgLeuGluHis 20
Db 4628 ATGCGCTCGCTGAAGCGCGCTTGGCGGCTTGCAGATCGCTTCTCGATCCGCTGTCGAC 4687
Oy 21 TTPAlaLeuThrArgProGluGlnThrCysValAlaAlaArgAlaAlaAngIyGluTrp 40
Db 4688 TGGCGCGCGTGGCGCCCGAGACACTTCATTGCGCGCCGCGAGCGCGTGGCGACTGG 4747
Oy 41 ArgArgIleSerTyrrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 4748 CGCGTGTGCTGATGAGCGATGCTCGACAGCGTGGCGCCATCGCCCGCGGCTGCTT 4807
Oy 61 ProTyrrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 4808 CGTTACGGCTGTGCGCCGACAGCGCGCTGCTGCTGCGGCATGACATCGAACAC 4867
Oy 81 LeuGlnLeuAlaPheGlyAlaMetTyrrAlaGlyIleProTyrrCysProValSerProAla 100
Db 4868 CTGCAACTGGCGCTGGCGCGGATGTACGCGGGATTCCTATTGCGCGGTCGCGCGCC 4927
Oy 101 TyrrSerLeuLeuSerGlnAspLeuAlaValLeuArgHisIleValGlyLeuLeuGlnPro 120
Db 4928 TATTCTGCTGTGCCAGGACTTGGCCAGCTGCGCATGTGTGTCGCACTGTGTGCAAGCCG 4987
Oy 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
Db 4988 GAGCTGATGTTCTGACGCGCCAGCGCGCTTACAGCGGCGCATTCGCGCGGTGTCGCG 5047
Oy 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPhe 160
Db 5048 CCCGAGACCGCGGTGATCAGCGTGGCGGAGGTGCGCGGCGCGCGCGAGCGGAGTTC 5107
Oy 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
Db 5108 GCCAGCTGTGGCCAGCGCGGTGGCGCGGAGCGCGCATCGCGCTTTCGCGCACCGGCG 5167
Oy 181 ProArgThrIleAlaValPheLeuPheLeuPheThrSerGlySerThrIleValLeuProVal 200
Db 5168 CTGAGAGCATCCCAAGTTTCTTCACTCGGAGCTGACCAAGCTGCGCCCAAGCGCGTG 5227
Oy 201 ProThrThrGlnArgMetLeuCyValAsnGlnGlnMetLeuLeuGlnThrPheProVal 220
Db 5228 ATCAGCACCAGCGCATGCTCTGCGCAACAGCAGATGCTGCTGAGCACTTTCGCGTG 5287
Oy 221 PheGlyGlyGluTrpProValLeuValAspTrpLeuProTrpPheHisIleThrPheGlyGly 240
Db 5288 TTTCGCTGAAGAGCGCGGTGCTGTGTGACTGCTGCGTGGAGCAACACTTTCGCGGCG 5347
Oy 241 SerHisAsnIleGlyIleValLeuTyrrAsnGlyGlyThrTyrrTyrrLeuAspArgGlyLys 260

```

```

Db 5348 AGCCATTAAGCTGGCATTCGCTCTACACGCGGATCTTTTACCTGAGCGAGCGGACG 5407
Oy 261 ProThrAlaGlnIyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
Db 5408 CCCAGCGCCCAAGCTTCGCTGAACCTTCGCGCAACTCAAGAGATTTCCCCACCGCC 5467
Oy 281 TyrrLeuThrValProIleGlyTrpGluGluValGlyAlaLeuGluTrpArgAspSerThr 300
Db 5468 TACCTAGCGCTGCGCCAGAGCTGAGAGATGTGTCAAGCCTTGAGAGAGCGCGAG 5527
Oy 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSer 320
Db 5528 CTGCGGAGATGTTTTTCAAGCGCATGAGCTGTTTTTCTTGGCGCGCGGCTTTCG 5587
Oy 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGlyAlaGlyIleArg 340
Db 5588 CAAGATGCTGGGATGCGCTGGAACCGGATCGCGACAGCATTTGCGCGCAACGATCCGC 5647
Oy 341 MetMetAlaGlyLeuGluGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360
Db 5648 ATGATGCGCGGCTGCGCATGACCGAGCGCGCGCTTTCGACCTTCACACCGAGACCG 5707
Oy 361 LeuSerMetAlaGlyTyrrIleGlyLeuProAlaProGlyCysGluValLysLeuValPro 380
Db 5708 CTGTGATGCGCGGTTTACATCGGCTGCGGCGCGCGGTTTGGAGATGCGCGTGTGCG 5767
Oy 381 ValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrrTrpArg 400
Db 5768 GTGATGCGCAAGCTCAAGAGCGCTTTTCGCGGCGCCACATCATGCGCGGCTCTGCGCG 5827
Oy 401 AlaProGluGlnAsnAlaGlnAlaPheAspArgGluGlyTyrrTyrrCysSerGlyAspAla 420
Db 5828 GCGCGGAGAGACCGCGCAGAGTGTTCAGCGCCAGGCTTCTACTGCTCTCGGAGATGCG 5887
Oy 421 IleValLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspArgIleAla 440
Db 5888 CTCAGCTGCGCGACCGCGCGCGACCGCGCACTGGGCTGATGTTCCAGCGGCGCATCGCC 5947
Oy 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 5948 GAGGACTTCAGATGTGTGTCGCGGAGTGTCTCAGCGTGTGCGACCGCTGCGCAATCGCG 6007
Oy 461 ValLeuGluGlyGlySerTyrrValLeuAspValValAlaAlaProAspArgGluCys 480
Db 6008 GTCTCTGAAGCGCGCGCTTATGTTCAGGACTGTGTGTGCGCGCCCGCGACCGCGAGTGC 6067
Oy 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
Db 6068 CTGGGGCGCTGTGTTCCCGCGCTGTATGATGCGCGCGCGCTGGCGGGGCTGGGCGCC 6127
Oy 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
Db 6128 GAGCGCAGGAGCGCCAGAGTGTGGCCAGCGCGCGCTGGCGCAAGTGTTCGCGCACTGG 6187
Oy 521 LeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetThrValGlyLeu 540
Db 6188 TTTCAGTGTCTGATGTAGGCGCAGCGGAGATGCAAGCGCTGAGAGTGTAGCGGCTG 6247
Oy 541 LeuAspThrProProSerIleAspLysGlyLysValAlaThrAspLysGlySerIleAsnGln 560
Db 6248 CAGTGGAGCGCGCGCTCATCGACCGTGGGAGATCACGCAAGAGGCTGATCATTCAG 6307
Oy 561 ArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaLeuTyrrArgGlyLysAspGln 580
Db 6308 CGCGCGGCTGTGAGTGGCGCGCGCAGAGTGTGAGAACTGTATCGCGCGCGCGAGCGG 6367
Oy 581 SerMetLeuArgAspGlu 586
Db 6368 TCGATCTGCGCGCGAG 6385

```

RESULT 6
AE016786/c

LOCUS AE016786 301214 bp DNA linear BCT 12-DBC-2002
DEFINITION Pseudomonas putida KT2440 section 13 of 21 of the complete genome.
ACCESSION AE016786 AE015451
VERSION AE016786.1 GI:26557030
KEYWORDS
SOURCE Pseudomonas putida KT2440
ORGANISM Pseudomonas putida KT2440
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 301214)
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
Beanan,M., Deboy,R., Daugherty,S., Kolonay,J., Madupu,R.,
Nelson,M., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
Holzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
Rizzo,M., Lee,K., Kosack,D., Mostl,D., Medler,H., Lauber,J.,
Hohstetel,J., Stracz,M., Heim,S., Kiewitz,C., Eissen,J., Timmis,K.,
Dueterhoff,A., Tummeler,B. and Frazer,C.
Complete genome sequence and comparative analysis of the
metabolically versatile Pseudomonas putida KT2440
Environ. Microbiol. 4 (12), 799-808 (2002)
2 (bases 1 to 301214)
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
Beanan,M., Deboy,R., Daugherty,S., Kolonay,J., Madupu,R.,
Nelson,M., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
Holzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
Rizzo,M., Lee,K., Kosack,D., Mostl,D., Medler,H., Lauber,J.,
Hohstetel,J., Stracz,M., Heim,S., Kiewitz,C., Eissen,J., Timmis,K.,
Dueterhoff,A., Tummeler,B. and Frazer,C.
Direct Submision
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA
FEATURES
source
1..301214
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
complement(69..1007)
/locus_tag="PP3196"
/note="similar to GB:X79440, SP:Q16798, and PID:535012;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="nodulation protein d1"
/protein_id="AA068804.1"
/db_xref="GI:24984828"
/translation="MRFHLDLNLVALDILLEQNITRAERLHMTQSGVGLRL
RPFPEELIVQVQKOPTPYALAEKAPVREVLITRASITAKPVDPAPASHPVLA
ASDYLISVFAOVYOKIHOEAPQITFEMIGPQDSSEMLRGEVDMYIPERYLHGF
PAQLFEEDHVCVMKRDNPVEVSHLTLEQYMOGHISVFGKRRHLISIDMTNNQGF
NRRLEVTINDENTLPLQLLVGYRVATMRRLRLVAGYPLRLPLPPVKI PVMREPL
MHRSMQDPMHRLRERISDVIOCASTQDALRATA"
1284..2225
/locus_tag="PP3197"
1284..2225
/locus_tag="PP3197"
/note="similar to GB:X54232, SP:P35052, and PID:31847;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glyoxalase family protein"
/protein_id="AA068805.1"
/db_xref="GI:24984829"
/translation="NNILGLDLVFGVDIOACADCLRDYGDVAOVLDQDGRFEALD
GSIITIRGDDGLPPALPPTPSLRETVYVAGSSTDLAIADLIGDRPVRDASGLV
HHPDIFPAIAFOVCRBPBPAADI.TNAPGAPORPINOGLIAPMPALPTLSLVV
YVPPDAKGEAFYRDLRGFTVTSDFGVGPPMRTAMDHNCIFMIOTPPHQGCHEF
THMGSGTEVLLAGTRFOOQWTSFVGPRHLIGSWFWYFNSPLGCHLEYDADMDKH
DMAARQAPMSADNSQLFLPASREKMAPSGPAKQG"

gene 2238..2576
/locus_tag="PP3198"
2238..2576
/locus_tag="PP3198"
/note="identified by match to PRM protein family HMM
PF00355"
/codon_start=1
/transl_table=11
/product="ferredoxin, 2Fe-2S"
/protein_id="AA068806.1"
/db_xref="GI:24984830"
/translation="WELCLDELGEGQARGFPOGRGADSLFALRYLGQVRYRNLCP
HMLVPLEYRKNNFLSADQWIVCAHGARFPEDETCVHGPGRGALELDHKEAGW
LLPLPEQDR"
2716..3840
/locus_tag="PP3199"
2716..3840
/locus_tag="PP3199"
/note="similar to SP:P23262; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="monooxygenase, putative"
/protein_id="AA068807.1"
/db_xref="GI:24984831"
/translation="NNTVNVLIIVGGIGGLCAIATLRKGIADVLDVELKTEMTYGV
GIIQSNVYREHRLGLEAYADAAVAFEDVAIYTTGEQQLARIQOGLAGQVAVN
GISRRALKVSETAIEIGTOYRLGTSVESPEQDSADVDFPTDSQGRGLMGVADG
LFSKRSVLFCRGYOPRFTGQSVWRYNPPRAGADHILANQCPSCNAGIPLADLMY
LFTTSHPEGNPMEPFPALADGLRQLAGFGGLIGLREQIVNAGVYKPLAVATYD
RWYCGRVLLIGDAHAATTPLQAGAGM1EDVAVLAELTACGSLDAOLQRMGRFPE
RCKFISEMSLLGVKEIQRPSPFRIGLVKMLVTAQPI"
3885..5597
/locus_tag="PP3200"
3885..5597
/locus_tag="PP3200"
/note="similar to GP:3184126, and GP:3184126; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA068808.1"
/db_xref="GI:24984832"
/translation="MPTPTHNKKEETPMLTRKGAFAFPLPALKPLSLVALAASQA
QAFEDTGNPDLRVWNTLKISAAARTODASSKTEGOTALNLDDGBNPFKGLISN
RDLTSLSDITVNVNGARVSGAAYDDVYNGDTNNDPRANYSVDYDAFPDTRKL
HGRKEILDARIPGKTEIGDMPLSGRLQVAMQGESLPEFGNNGIAGGAAPDVVXGL
SVPTQFKELRIPVQIISGQLITPDVSVYGAITYOREANRLPAAGSYFTSTDFPQDS
AERMPVGAVPFPGQPLAFPHGNDKQAKDSQGGQVRRRTETVDMGLYAIRFHDKSP
QLNVRPNPANDLPVSGKAGEFYWVYPEGIEAGFASFTVGNFNLAGELSTWNPPLA
STNQHAAVGEA1IDNDNPVYATGRTLHANVSLASLEPNFYAESSTFGEIATMNRVM
SVSKNADVPADSDATSLRVFEPWROLPGDLSPVIGASVTNGSMALGTFG
TDHGQDNLVIGLGNVLTNWSLGLYTHHYGPENTFLDADNHTFPQSLKDRFLAFTV
SRTP"
7034..8095
/locus_tag="PP3201"
7034..8095
/locus_tag="PP3201"
/note="identified by match to PRM protein family HMM
PF02012"
/codon_start=1
/transl_table=11
/product="BNR domain protein"
/protein_id="AA068809.1"
/db_xref="GI:24984833"
/translation="NNKRLYGMWLLAACGCLPPAFAGDSYDVLAQPALHPQALRAVLO
DVAAADARIVAVGEGVALLSDNGATWQAQAAVAVGVSLTAVQF1IDANGAVAGAG
VTLHQQDGEHWSVOLDGRRAAALDEVQAAEAADRRRLAAARLADADKPLALNF
TDARGLVGAAGVAMRTDDGGRSMQSEIGRLPNRGRGLYVALAQGVYTAGQGL
ILRSDDGHEHYALGPYDGSFAATVLPDGLLIGRGLSFASDDGQVYTAGQLANP
TPASINGRIVAGQULLANQGMRLRSGIGAFAPAAQPAVAVSDPLPLTAATTADSGIVA
VGAQARRLEBSLSTTD"
8108..10588

Db 169647 CGCTCCCGCAGACCGCGCGCTTCGACGAGAGGGCTTCTACTGTTCCGGCGAC 169588
Qy 420 AAla1LysLeuAlaAspProAlaAspProGlnLysLeuMetPheAspGlyArgIle 439
Db 169587 GCGTTGAGCTGGCCGATGCGACGACCGCGAGCTGGCTGATGTTGATGGCGGATC 169528
Qy 440 AAlaGluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArg 459
Db 169527 GCTGAGACTTCAAACTTTCGTCGGGGTATTCGTAGTGTCCGGCCGCTGCCAACCCG 169468
Qy 460 AAlaValLeuGlnGlyGlySerTyrValLeuAspValValAlaAlaPheAspArgGlu 479
Db 169467 GCAGCTCGAGGGCTCGCGCTTACGACAGACATGTTGTCACCGCGCGACCGGAA 169408
Qy 480 CyAlaLeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGly 499
Db 169407 TGCCTGGGCTGCTGTTGTTCCCGCTCTCCGAGTGTGGCGCGCTGGCGGCTGGCA 169348
Qy 500 LysGluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAsp 519
Db 169347 GAGAGTGCACGCGATGCGGGGTGCTGGCCACGACCGCTGGCAGTTGTTCCGTGAC 169288
Qy 520 TrpLeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGly 539
Db 169287 TGGCTGAGGCTTGAACCGCGATGCCCAAGGACGCGCGCTGATCGAATGGCTGTCG 169228
Qy 540 LeuLeuAspThrProProSerIleAspLysGlyGlyValThrAspLysGlySerIleAsn 559
Db 169227 CTGCTGCCGACGCGCGCTGATCGAGCGCGGTGAAATCCCGACAAAGGCTGATCAAT 169168
Qy 560 GlnArgAlaValLeuGlnTrpArgSerAlaValValAspAlaLeuTyrArgGlyGluAsp 579
Db 169167 CAGCGCGCGCTGTCAGCGCGCGCTGAGTCAAGTCAAGCGCTGTAACGTGCGAAGAC 169108
Qy 580 GlnSerMetLeu 583
Db 169107 CCCGACGCAATTG 169096

RESULT 7
PPUS36324 4407 bp DNA linear BCT 09-JAN-2004
LOCUS Pseudomonas putida ech gene, vdh gene, fcs gene and aat gene
DEFINITION (partial).
ACCESSION AJ536324
VERSION AJ536324.1 GI:40788111
KEYWORDS aat gene; beta-ketoliase; ech gene; fcs gene; feruloyl-CoA hydratase; feruloyl-CoA synthetase; vanillin dehydrogenase; vdh gene.

SOURCE
ORGANISM Pseudomonas putida
Pseudomonas putida
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1
REFERENCE
AUTHORS Ficca, A.G., Di Gioia, D., Barghini, P., Fava, F., and Ruzzi, M.
TITLE Identification of Pseudomonas putida strain T2 genes involved in ferulate catabolism
JOURNAL Unpublished
AUTHORS Ruzzi, M.
TITLE Direct Submission
SUBMITTED (09-JAN-2003) Ruzzi M., Agrobiology and Agrochemistry,
DABAC - University of Tuscia, via C. de Lellis, snc - Viterbo,
01100, ITALY
FEATURES
source
1..4407
Location/Qualifiers
/organism="Pseudomonas putida"
/mol_type="genomic DNA"
/strain="T2"
/isolation_source="soil"
/db_xref="taxon:303"
/country="Italy:Latvio"
10..741

CDS
/gene="ech"
10..741
/gene="ech"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="feruloyl-CoA hydratase"
/protein_id="CAD60261.1"
/db_xref="GI:40788112"
/translation="MSPTLNRMRDYLLETVEODADARVLVLNGESAMTAGMDLKEYF
REVDAPEILQRIIRDASEMOMKLLRFPYSKPTIMVNGMCGGSGSPILVADILCA
DEATRGSLSTINGITPPGNIVSKAMADTVGHRSLIYITMGKTFDQKAOMGIVKYSV
PLAQRLDEVLLAODLIDNPVLLRAKNGFRCELTEWQEDYLYALQDAELDPE
HRAQGLKQFLDLSKIKPGLQAYKR"
782..2239
/gene="vdh"
782..2239
/gene="vdh"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="vanillin dehydrogenase"
/protein_id="CAD60262.1"
/db_xref="GI:40788113"
/translation="MDVKGALLIGKAPASDGTFFERRHPVSGEVSVVAATLTAD
ADAAYEAQAAPMAALQPGERRRLAKAALALARPDEFLAMGETAKANWGFN
VMAANIILEASMTTOIQGSVIPSDVPSFAMALROPCEVLIAFMNAPVILATRA
IAMPACGNTVVLKASEAPVHGLIGVLAAGLGDGVNVIYCNAPHPADAVVERLI
AHPAYRVNFTSGTIVGQLAARHLKPAVLEIGGKAPLVLAADADLEAAVADAA
PBAYRNQGIOMSTERLIVDQAVABEPFAKLEKRYRLAADPSQDAPALGSLIDGA
GRIQALVDAVGKATLVAGSERGTSVMQPVLDHVNAGMKLIRSESGPVAAVLRGA
SGDEALALANDSEFGLSAIFSRDQRLALAQRLSEGIHINGPTVHDEAQMPGG
VVASGSGFSGTAIDAFQLRWVTVQHGARRHYPI"
2319..4175
/gene="fcs"
2319..4175
/gene="fcs"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="feruloyl-CoA synthetase"
/protein_id="CAD60263.1"
/db_xref="GI:40788114"
/translation="MDCELTLSVAGRAPLRAVKGIPQPVQVRQGGVLYLDPVETLDAY
PARLMERLHMAQORSEQTFVARADGQMCHITTAAMQORVRLATWLGGLSHER
PLIISGNDLEHVOATALVAGIYPCVPSPAYSLSDPGRHIVOTLQGLVFAA
DAAYRAIDAVEPESINIVLGGELAGASPTPAFLVTTVAQPCADQAPALRPDLSA
KLPITSGSTKLPAKAVTTQGMICANQOMLQTPPEPAAVPPVINDMLPNNHFGSGHN
VGIVLVNGSTLVIDGKPTPOGFAQTLNHLKDISPAVLTVPKMBELVKALBODBAL
CERFFQRMKLPFFPAGAGLSQALWDRLDAAVAHCGERIRMAAGLQMTETSPCTPTG
PLSVAGYIGLPAPGCEVYKLVYVDGKLEAFRGPVHWPGYMRSSQOQBAFDEGFCG
GDAVNLCDPENGQILMPGPRIAQDFKLSGVFVSPWRTVSSGSPYLVVVVYGP
DRECIQILFPLRPECROLSGLGAQSPFKVLASPVPRCLASLATLRGATGNASV
IRPAGLAAPRPIDRGETIDKSLNQRVQLRHGLIDSLVAGPELRVDAE"
4172..4407
/gene="aat"
4172..4407
/gene="aat"
/codon_start=1
/transl_table=11
/product="putative beta-ketoliase"
/protein_id="CAD60264.1"
/db_xref="GI:40788115"
/translation="MTARGQFRTPAATALLGRIPTPWVNDINGAFGLVSPFDLIGIRAG
AVLQARLADGSDVSLAGSMQADFDALILPRH"

ORIGIN

Alignment Scores:
Pred. No.: 2..56e-140 Length: 4407
Score: 2103.00 Matches: 410
Percent Similarity: 79.79% Conservative: 48
Best Local Similarity: 71.43% Mismatches: 114
Query Match: 68.23% Indels: 2

DB:	1	Gaps:	2
US-09-750-986D-30 (1-589) x	EPUS36324 (1-4407)		
QY	4	LeuGIuLaLeuLeuProPheProGIyAArgIleLeuGIuAArgLeuGlnIeTPRAlaLys	23
Db	2430	GTGGAAACGCTCGATGCCCTACCCGGCAAGCGCTGATGGAGCCCTTGCTCATTTGGGCGCAG	2489
QY	24	ThiAArgProGIuGIuThrCysValaAlaAlaArgAlaAlaAsnGIyGIuTPRArgIle	43
Db	2490	CAGGCGAGCGAGCAGACGTTGTGCGCGCGCGCTCGCGAGCGGCAATGGCAGACACATC	2549
QY	44	SerTyrLaIaGIuMetPheIleAsnValaArgAlaIleAlaGlnSerLeuLeuProTyrGIy	63
Db	2350	ACCTACCGCGCGCATGTGGCAAGCGGTGTGACGAGATGGCAAGTGGCTGTGCGCCMAAGC	2609
QY	64	LeuSerAlaGIuAArgProLeuLeuIleValSerGIyAsnAPLeuGlnIeLeuGIuLeu	83
Db	2610	CTGAGCGCGCAACCGCCCTATCTATCTCTCGGGCAAGCATTTGAGCAAGTGCATG	2665
QY	84	AlAPheGIyAlaMetTyrAlaGIyIleProTyrCysProValSerProAlaTyrSerIleu	103
Db	2670	GCCACGCGCGCGCTATGCGGATATCTCATTTCCCGGATGCGCGCATATTCAGT	2729
QY	104	LeuSerGIuAPLeuAlaLysLeuArgIleAlaValGIyLeuLeuGlnProGIyLeuVal	123
Db	2730	TTGTCCACGAGATTGTGGCAAGCTGGCATATCTGCAGACATGGCAACCGAGCTGGTC	2789
QY	124	PheAlaAlaAPAlaAlaProPheGlnArgAlaIleGIuThrIleLeuProAspArgVal	143
Db	2790	TTCCGCGGTATGCCGCC--TACGGCAGGGCATCGATGCGGTGTGAAACCTTCGATT	2846
QY	144	ProAlaIlePheThrArgGIyGIuLeuAlaGIyAArgThrValSerPheAspSerIleu	163
Db	2847	AACTGGTGTGGGCAAGGCGAGCTGGCGGGTGGGAGAGACGATTCGCGAGGATA	2906
QY	164	LeuGIuGlnProGIyGIyIleGlnAlaAspAsnAlaPheAlaIleThrGIyProAspThr	183
Db	2907	CTGACCAACCGTGGCCCAACCTGTGGCCACCAAGGCTTTTGGCGCTTGGCGCGGACAGC	2966
QY	184	IleAlaLysPheLeuPheThrSerGIySerThrLysLeuProValAlaValProThrThr	203
Db	2967	CTGGCAAGTCTCTGTACCTTCGGCTCCACCAAGCTGCCCAAGGCGGTGTACCAAC	3026
QY	204	GlnAArgMetLeuCysAlaAsnGIuMetLeuLeuGlnThrPheProValPheGIyGIu	223
Db	3027	CAGGCGATGCTGTGGCCCAACCAAGAGATGCTGTGGACACTTCCCGAGTTCGCGAA	3086
QY	224	GluProProValLeuValAspTrieuProTrrPheAsnIsthrPheGIyIleSerIleAsn	243
Db	3087	GTGCCGCGCGGTGTGTGACTGGCTGCCGTGGAAACACAGTTTGGCGAGCCACAAC	3146
QY	244	IleGIyIleValLeuTyrAsnGIyGIyThrTyrTyrLeuAspArgGIyLysProThrAla	263
Db	3147	GTGGGTATCGGTGTACACAGCGCGGCGCGCTGTACATCGAGATGGGCAACCGCAGCGG	3206
QY	264	GlnGIyPheAlaGIuThrLeuArgAsnLeuSerGIyIleSerProThrAlaTyrLeuThr	283
Db	3207	CAGGCGTTCGCCACAGCGCTGCACAACTCAAGAGCATTTCCCGACGGCTACTGAC	3266
QY	284	ValProLysGIyTPRGIuLeuValGIyAlaLeuGIuAArgSerThrLeuArgGIu	303
Db	3267	GTGCCCAAAAGCTGGAGGAATGGTCAAGGACATGAGACAGAACCGCGCCTGTGGAG	3326
QY	304	ArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGIyLeuSerGIuIle	323
Db	3327	CGGTTCTTCCAGCGCATAGCTGTCTTTTTCGCGGTGCCGGGTGTTCACAGCCATC	3386
QY	324	TyrAspArgLeuAspArgValAlaGIuGlnIsthrCysGIyGIuArgIleAArgMetMetAla	343
Db	3387	TGGAGCGCGCTTGAAGCGCGGTGGCGGTGGCCCACTCGCGGAAAGCATTCGCATATAGGCG	3446
QY	344	GIyLeuGIuMetThrGIuThrAlaProSerCysThrPheThrThrGIyProLeuSerMet	363

Db	Accession	LOCUS	DEFINITION	VERSION	KEYWORDS	ORGANISM	REFERENCE	Authors	Title	Journal	Medline	Pubmed	Reference	Authors	Title
Db	3447	GGGCTGGGCAATGACCGAGACTTCCTCCGTCGTGACCTTACACACCGGGCGCTGTCGGTGG	189050 bp DNA linear BCT 11-APR-2003	Al646077	Al646077	Ralstonia solanacearum	1	Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangenot, S., Arlat, M., Billault, A., Brodtier, P., Camus, J.C., Caticolico, L., Charlet, M., Choisme, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weisenbach, J. and Boucher, C.A.	Genome sequence of the plant pathogen Ralstonia solanacearum	Nature	415 (6871), 497-502 (2002)	2 (basee 1 to 189050)	Boucher, C.A.	Direct Submission	
Qy	364	AlaGlyTyrIleGlyLeuProAlaProGlyCysGluValLysLeuValProValAspGly	189050 bp DNA linear BCT 11-APR-2003	Al646077	Al646077	Ralstonia solanacearum	1	Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangenot, S., Arlat, M., Billault, A., Brodtier, P., Camus, J.C., Caticolico, L., Charlet, M., Choisme, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weisenbach, J. and Boucher, C.A.	Genome sequence of the plant pathogen Ralstonia solanacearum	Nature	415 (6871), 497-502 (2002)	2 (basee 1 to 189050)	Boucher, C.A.	Direct Submission	
Db	3507	GCCGGCTACATCGGCGCTGCGCGCGCGGGGGCGAGGCTCAAGCTGAGTCCGGTGCAGCGG	189050 bp DNA linear BCT 11-APR-2003	Al646077	Al646077	Ralstonia solanacearum	1	Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangenot, S., Arlat, M., Billault, A., Brodtier, P., Camus, J.C., Caticolico, L., Charlet, M., Choisme, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weisenbach, J. and Boucher, C.A.	Genome sequence of the plant pathogen Ralstonia solanacearum	Nature	415 (6871), 497-502 (2002)	2 (basee 1 to 189050)	Boucher, C.A.	Direct Submission	
Qy	384	LysLeuLugLysArgPheHisLagGlyProHisValMetSerGlyTyrTTPPAAGAlaProGlu	189050 bp DNA linear BCT 11-APR-2003	Al646077	Al646077	Ralstonia solanacearum	1	Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangenot, S., Arlat, M., Billault, A., Brodtier, P., Camus, J.C., Caticolico, L., Charlet, M., Choisme, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weisenbach, J. and Boucher, C.A.	Genome sequence of the plant pathogen Ralstonia solanacearum	Nature	415 (6871), 497-502 (2002)	2 (basee 1 to 189050)	Boucher, C.A.	Direct Submission	
Db	3567	AAAGCTCAGAGCGAGATTTCGGGGGCCCCACAGTGAATGCGGGGATCAAGCGCTATCCAG	189050 bp DNA linear BCT 11-APR-2003	Al646077	Al646077	Ralstonia solanacearum	1	Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangenot, S., Arlat, M., Billault, A., Brodtier, P., Camus, J.C., Caticolico, L., Charlet, M., Choisme, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weisenbach, J. and Boucher, C.A.	Genome sequence of the plant pathogen Ralstonia solanacearum	Nature	415 (6871), 497-502 (2002)	2 (basee 1 to 189050)	Boucher, C.A.	Direct Submission	
Qy	404	GlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleLysLeu	189050 bp DNA linear BCT 11-APR-2003	Al646077	Al646077	Ralstonia solanacearum	1	Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangenot, S., Arlat, M., Billault, A., Brodtier, P., Camus, J.C., Caticolico, L., Charlet, M., Choisme, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weisenbach, J. and Boucher, C.A.	Genome sequence of the plant pathogen Ralstonia solanacearum	Nature	415 (6871), 497-502 (2002)	2 (basee 1 to 189050)	Boucher, C.A.	Direct Submission	
Db	3527	CAGACCCAGAGAGGCGCTTCGACGAGAGAGGGGTTCTACTGACGGGTGATCGCGTCCGCTG	189050 bp DNA linear BCT 11-APR-2003	Al646077	Al646077	Ralstonia solanacearum	1	Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangenot, S., Arlat, M., Billault, A., Brodtier, P., Camus, J.C., Caticolico, L., Charlet, M., Choisme, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weisenbach, J. and Boucher, C.A.	Genome sequence of the plant pathogen Ralstonia solanacearum	Nature	415 (6871), 497-502 (2002)	2 (basee 1 to 189050)	Boucher, C.A.	Direct Submission	
Qy	424	AlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaLysPhe	189050 bp DNA linear BCT 11-APR-2003	Al646077	Al646077	Ralstonia solanacearum	1	Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangenot, S., Arlat, M., Billault, A., Brodtier, P., Camus, J.C., Caticolico, L., Charlet, M., Choisme, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weisenbach, J. and Boucher, C.A.	Genome sequence of the plant pathogen Ralstonia solanacearum	Nature	415 (6871), 497-502 (2002)	2 (basee 1 to 189050)	Boucher, C.A.	Direct Submission	
Db	3687	TGCAGACCCCGGCAACCCGACAGCTGGGTTGATGTTTCAGCGGGCGCATTTGCCAGAGATTCC	189050 bp DNA linear BCT 11-APR-2003	Al646077	Al646077	Ralstonia solanacearum	1	Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangenot, S., Arlat, M., Billault, A., Brodtier, P., Camus, J.C., Caticolico, L., Charlet, M., Choisme, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weisenbach, J. and Boucher, C.A.	Genome sequence of the plant pathogen Ralstonia solanacearum	Nature	415 (6871), 497-502 (2002)	2 (basee 1 to 189050)	Boucher, C.A.	Direct Submission	
Qy	444	LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu	189050 bp DNA linear BCT 11-APR-2003	Al646077	Al646077	Ralstonia solanacearum	1	Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangenot, S., Arlat, M., Billault, A., Brodtier, P., Camus, J.C., Caticolico, L., Charlet, M., Choisme, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weisenbach, J. and Boucher, C.A.	Genome sequence of the plant pathogen Ralstonia solanacearum	Nature	415 (6871), 497-502 (2002)	2 (basee 1 to 189050)	Boucher, C.A.	Direct Submission	
Db	3747	AAAGTTGTCCTCCGGGGGTGTTCTGACGGTTGGCGCATGCGCACCCAG--GTAATCTTCA	189050 bp DNA linear BCT 11-APR-2003	Al646077	Al646077	Ralstonia solanacearum	1	Salanoubat, M., Genin, S., Artiguenave, F., Guzy, J., Mangenot, S., Arlat, M., Billault, A., Brodtier, P., Camus, J.C., C							

JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Bodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT
http://sequence.toulouse.inra.fr/R.solanacearum.html.
location/Qualifiers
1. 189050
/organism="Ralstonia solanacearum"
/mol_type="genomic DNA"
/strain="GMI1000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
79. 1500
/note="R49+ or RS06146
LONG DIRECT REPEAT 49 +
predicted by Homology"
/evidence=not_experimental
83. 2971
/gene="RSP0175"
/note="synonym: RS04687"
83. 2971
/gene="RSP0175"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE VGR-RELATED PROTEIN"
/protein_id="CAD17326.1"
/db_xref="GI:17430643"
/db_xref="GOA:Q8XTD8"
/db_xref="TrEMBL:Q8XTD8"
/translation="MDFKTLQHLFAPARHLYALBEGCPFIABLVAEWLIGREALSELF
EMRVASANARFALSPFGORVTLVTLADGQARGLIRAOELGADGSLARYL
TVPMFWLATOQSHSOVQPNRPLADLLEOVLSPYAPAMRPAAGADEMMAARGTSH
IAOFRTDPRFTRRLAELAGCTTTVEDQAPSGHLLIFADSPRLAEDTASABEGI
RTHRASQERDAIQALICHSRTAVGVVAAMDQAQKAFRAHVPSRCGIASPDA
YLSISPLAPDVTANAQRIAEQVESVEARARLFIGSVRLTASGRLKIVDCPLPK
DVDPYPLLDIVHEGGINLLTADQTATLAKLGIDALITPDTPSPSPESGPGPFG
MAPHEVIERFTPADLLTARAHGAVGQFRAGDARPMRPVSHPTGRLVAEPTAG
VGSATVVGNGETFEASDAEHHTTPRGQVVRPMQGRPDORSTGMILVQROGA
GNGWOLPRIGQEVLYKPOEDDIDQPVVYGALYNGGEGAIPTPGQVAKYRQAD
PEFLYLSGDSASAOGNLAHSPAMHGLTEABGRNNAALSGFKSAHGGRANL
LVFDDSDQLRQLATQAYSQLNLGHLHQDNRRGSPRGQFELRTGQYAVRQQA
GVLLTYRDATSGKAVPTGDNAAGILIRQAKQLTSLTSGATHTQTALSTPAKQDA
PLAKOKAASGWIDKALDAQODAAAGNATQGVKPVHSEATVVOLAGAAYVIAIGQ
DLOFANGESVULASGODTNVAVSKARIHAGGIGVAGLSTAGDNDNIGLQITAGQDN
LDVQAOHDALGLISKDDKLVSKNLHVDPAAARIRLATAGASTLLEGNTITVEPTG
RITVKAQRSFAPVTSYPLVPFGQVCECVLRMGKOHAFANNG"
1501. 2865
/note="R650+ or RS06148
PROBABLE LONG DIRECT REPEAT 50 +
predicted by Homology"
/evidence=not_experimental
2999. 3835
/gene="RSP0176"
/note="synonym: RS04688"
2999. 3835
/gene="RSP0176"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical

predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD17327.1"
/db_xref="GI:17430644"
/db_xref="GOA:Q8XTD7"
/db_xref="TrEMBL:Q8XTD7"
/translation="MATTRQIQDPHRRPRAWVDYPADAVEQLLSALPEPLDYALVDN
AYDPARHRRPRPMPRSLXAGRYGQGLDAPLTPMRLPDAVABERTGLDPLLO
ETSGKPMISFTLHANAADPLTTLNQNQEAHVHGETAFILRLADTNALDAVQVRDA
QORFLSDIQWCFRRDSLRHSHAYGGGGAESTPYKCTAQVQLEALRPGAML
NIVQNSPHIFGDLTGLPSQAYECIMALQALBPQNPHTDAVALRLIASLNRAGLQH
AA"
4412. 6634
/gene="RSP0177"
/note="synonym: RS04689"
4412. 6634
/gene="RSP0177"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD17328.1"
/db_xref="GI:17430645"
/db_xref="GOA:Q8XTD6"
/db_xref="TrEMBL:Q8XTD6"
/translation="NSVLEPQDPNPALTVDEQARFLACRMQEFERHCSTERHLSFF
DSTNNRRKYDTPROHSHSVARLFDIPEDEHGRIRYPERGIPFERKIDTGGRDAR
AGIAGWGQGEARINNALQVTNALAYYQTSLSAMGDELIVHQSSDKMLIGA
VTTAGRDEVEOLANAKNSEMYTAAETVYTPRHEBERVALLQREBEYSEKILALIGA
RKRMILCTRISVGFSGRGAEARVANNMKDLDDMTLAGVAFDFLGIPTVAVS
GVANSTKYATGSHSGGEARFLRIPGVYVRLVLSAHEHRSSEPLDKVANDCLEYA
PEVHTDVGAAYOPGQGRCCGTGKRPDSNKLSQTLTAMRYEAAAAGPLPMARNL
TQEVRAALKISPDLLQATNDYDVANPLIRKHGGTIGARVQYGLYLKMRWRLLGG
AQAERNOQLFFQABQYGAQCANDLSEVANALLKEAKDLAARENDAYDGMNQYQVR
LEVTOGVAVWNRQKWDGKYREVRVAYWNTAPLDPRLIVRLICDVIHDSRAMFK
PGASDQSVWNRQOARLEOLKOQAAWQVADISRLIGTLKYEESTANGDGEAA
EMPVVIGORRDLQOYLKXGQSVMEKGRSPSMGVIARWRTHRAPLPTLIGERAQAMD
EVASVPGKAKAQAADADAEARLANTARKLLEAQOBSLERRATTAQLQFLGGYVPR
"
6745. 7527
/gene="RSP0178"
/note="synonym: RS04690"
6745. 7527
/gene="RSP0178"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD17329.1"
/db_xref="GI:17430646"
/db_xref="GOA:Q8XTD5"
/db_xref="TrEMBL:Q8XTD5"
/translation="MTQYLYHTTTTAAVARIIKTKGLTLAAHPEALGRPYARRHGAFEV
NRAAQEPGROVRLKAYLKKGLGAGYSIDQIRTGORPFTPIVVPAGNRDDQVEITR
VBEAEVRLAIGTPANKPGRLTMDLTIGEHADMMLRTRRANMLCLAVHTVLEY
AIEBGMTSHVTFSPERASDCYSYTRQHGGAACSVLRVSRMAAALPLDDPSDFRA
VMTQRRILLPOQLEIWRAPSDVLTNADPRAAGNMPLTQWS"
7518. 8021

/gene="RSP0179"
/note="synonym: RS04691"
7518. 8021
/gene="RSP0179"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by FrameD
/codon_start=1
/evidence=not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD17330.1"
/db_xref="GI:17430647"
/db_xref="GDA:Q8XTD4"
/db_xref="TrEMBL:Q8XTD4"
/translation="MEIAPDRMGQPTMRVAGGAIALSALALAPALADARPGMQ
DEPARAVSAQGEYVSGTHIPLVLIHADPAMVAQSTVDVDRHNAI PKGSRLI
GKEIRVNAAREIIVTIGLOIPASGTRIDPPIQATMRDSAGVADLSFGALLGAMTS
EPTIVPH"

gene
complement (8066. 8488)
/gene="RSP0180"
/note="synonym: RS04692"
complement (8066. 8488)
/gene="RSP0180"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage

Alignment Scores:
Pred. No.: 1,39e-126 Length: 189050
Score: 1940.50 Matches: 373
Percent Similarity: 75.43% Conservative: 69
Best Local Similarity: 63.65% Mismatches: 143
Query Match: 62.96% Indels: 1
Gaps: 1

US-09-750-986D-30 (1-589) x AL646077 (1-189050)

Qy 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyValArgIleLeuGluArgLeuGluHis 20
Db 66491 CTGGCTCGGCTAGACCGGATCGGCTATCCGAGCGATGACGATTCGCTGTCGCC 66550
Qy 21 TTPAlaLysThrArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTyr 40
Db 66551 GGGGCGGACAGATCCGATCCGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 66610
Qy 41 ArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 66611 GAGGCGATCACTATGCGAGATCTCTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 66670
Qy 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 66671 GCGGCGGCGGCTGTCGCCGAGCGGCGGCTGCTGATCTCTGCGGCGGCGGCGGCGGCT 66730
Qy 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAla 100
Db 66731 CTGAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 66790
Qy 101 TyrSerLeuLeuSerGlnAspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnPro 120
Db 66791 TACCGCTGGTGAACGAGCACTACAGCGGCTCTCCCTACGCTGGGCGGCGGCGGCGG 66850
Qy 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGlnThrIleLeuPro 140
Db 66851 GGCTGGGCTATGCAACGAGCGGCTGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 66910
Qy 141 AspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyArgArgThrValSerPhe 160
Db 66911 CCGGAGACCGAGGCTCGGATGACACGAGGAGGTGGGCGGCGGCGGCGGCGGCGGCTG 66970

Qy 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaIleThrGly 180
Db 66971 GGCTCGGCTGCTC---GGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 67027
Qy 181 ProAspThrIleAlaLysPheLeuPheThrSerGlySerThrIleLeuProIleValAla 200
Db 67028 CCGGACAGATCCCGCAAGTTCCTCGAGCTCGGCGGCTCGACCGGCGGCGGCGGCGGCTC 67087
Qy 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnGlnMetLeuLeuGlnThrPheProVal 220
Db 67088 ACCACACGATGCAATGCTGTGACAGACACAGATGCTGTGACAGCTTTCCTGCG 67147
Qy 221 PheGlyGluGluProProValLeuValAspTyrLeuProTyrAsnHisThrPheGlyGly 240
Db 67148 TTGGGAGGCGGCGGCGGCGGCTGCTGAGCTGCTGCGGCGGCGGCGGCGGCGGCGG 67207
Qy 241 SerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspAspGlyLys 260
Db 67208 AGCCACACGCTGGGCTCGGCTGTACACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 67267
Qy 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
Db 67268 CCGACACCGGCGGCTTCGAGCAGAGCTGCGAACCTGCGGCGGCGGCGGCGGCGGCTG 67327
Qy 281 TyrLeuThrValProIysGlyTyrGluGluLeuValGlyAlaLeuGluArgAspSerThr 300
Db 67328 TACTTCAAGCTGCGCCAAAGGCTGGAGATCTGACGAGCGGCTGACGCGGCGGCGGCGG 67387
Qy 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSer 320
Db 67388 CTGCGGAGCACTTATGCGCGCGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 67447
Qy 321 GlnGlyIleThrAspArgLeuAspArgValAlaGlnGlnHisIleCysGlyGlyArgIleArg 340
Db 67448 CAGGCGCGGCGGAGCGGCTCGAGGCGGTGACCGAGGCCATGCGGCGGCGGCGGATCCG 67507
Qy 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360
Db 67508 ATCATGCGCGGCTGGGAGTGAACGAGACCGGCGGCTGTCGATGATCAACACCGGCGG 67567
Qy 361 LeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGluValLysLeuValPro 380
Db 67568 GTCATATGCGCGGCTGATCATCGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCTG 67627
Qy 381 ValAspGlyLysLeuGlyArgPheHisGlyProHisIleAlaMetSerGlyTyrTyrArg 400
Db 67628 GTGACCGGCAAGCTCGAAGCGGCTTCAGAGGCGGCGGCGGCGGCGGCGGCGGCTG 67687
Qy 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAla 420
Db 67688 GCGCGGAGCTGGCGGCTGTAAGCTTGAACGAGAGAGGCTTACACCGCTCGGCGGAGCGG 67747
Qy 421 IleLysLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAla 440
Db 67748 GTGGCTTCGTGACGAGCGGCGGCGGCGGAGATGCGGCTGATGTTCAGCGGCGGATG 67807
Qy 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 67808 GAGGACTTCAAGCTCAAGCTCCGACCTTGTGAGCGTGGGCGGCGGCGGCGGCGGATC 67867
Qy 461 ValLeuGluGlyLysTyrValLeuAspValValAlaAlaProAspArgLysCys 480
Db 67868 ATCAGCGAGGCGGCGGCTTACGCTGACGCGGTATCCCGGAGTGAACCGGAGACAC 67927
Qy 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
Db 67928 CTCAGCGTGTGGGTTCGCGCGGCTGAGACATGCGGCGGCTGCGGCGGCGGCGGCGG 67987
Qy 501 GluAlaSerAspAlaGluValIleAlaSerGluProValAlaArgAlaThrPheAlaAspTyr 520
Db 67988 GGCCTCAGCGGCGGCGGAGTGGCGACAGCGCTGCGGTGGGCGGCGGCTGCGGCTG 68047
Qy 521 LeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeu 540


```

Db      68048 CTGACGCGCTGACACCGCGCGCACCGGCTCGGCCAACCGGATCGCCCGCTGCGCTGG 68107
Qy      541 LeuAspThrProProSerIleAspIlyGluValThrAspIySerIleAsnGln 560
Db      66108 CTGAAAGAGCGCGCTGCTGCACACGCGAGGTCCACCAAGGCGCTGATCCAAACGAG 66167
Qy      561 ArgAlaValLeuGlnIrrPArGSerAlaIyAlaAspAlaLeuTyrrArgIyGluAspGln 580
Db      66168 CGGCGCGTGTCCCGCGCGCGCGCTGATCGAGCGATCTACGCGGCGCGGACGCGC 68227
Qy      581 SerMetLeuArgAspGlu 586
Db      68228 AAGGTATGCTGCGCGAG 68245

RESULT 9
AR319843
LOCUS   AR319843 1902 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 2393 from patent US 6562958.
ACCESSION AR319843
VERSION AR319843.1 GI:33700946
KEYWORDS
SOURCE  Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1902)
AUTHORS Breton,G. and Bush,D.
TITLE    Nucleic acid and amino acid sequences relating to Acinetobacter
          baumannii for diagnostics and therapeutics
          Patent: US 6562958-A 2393 13-MAY-2003;
          Location/Qualifiers
            1..1902
            /organism="unknown"
            /mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1..28e-108 Length: 1902
Score: 1652.50 Matches: 310
Percent Similarity: 70.57% Conservative: 100
Best Local Similarity: 53.36% Mismatches: 168
Query Match: 53.62% Indels: 3
Gaps: 3
US-09-750-986D-30 (1-589) x AR319843 (1-1902)
Qy      5 GluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluIleSTrPalAlaIySThr 24
Db      133 GAGCAATTAAACCGGATCCACAAACCTGACAGATCGTTTAATTCATTTTGACAAACC 192
Qy      25 ArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTrPArGArgIleSer 44
Db      193 AACCCGACATATTTTTCAGACAAAACGCAATGCTCAAGGCGAATGGCTCAACAGT 252
Qy      45 TyrrAlaGluMetPheHisAsnValArgAlaAlaAlaGlnSerLeuLeuProTyrrGlyLeu 64
Db      253 TATGAGAAGTGTCTACAAACGCGATGCGACATGCTCAAGCTTTGCTGCTGCTAATTTA 312
Qy      65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluIleLeuGlnLeuAla 84
Db      313 AGCGAAGAAAGACCTTGTCATTTTAAGGTATGATCTCGAACAATTTAAACACGCTCT 372
Qy      85 PheGlyAlaMetTyrrAlaGlyIleProTyrrCysProValSerProAlaTyrrSerLeuLeu 104
Db      373 ATGGCTGCCATGCTGCGACAGCGCTGCTTTCGCGCTAATTTCCCTGCTACCTCTGATT 432
Qy      105 SerGlnAspLeuAlaIySerLeuArgHisIleValGlyLeuLeuGlnIrrProGlyLeuAlaIyPhe 124
Db      433 TCTCAAGACTTTGGCAAACTCAAAACATGTGTTGAAGTGTCTCAACCTGATAGTTTAT 492
Qy      125 AlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValPro 144
Db      493 GCGAGGATGACAAAGCTTTTGCCAAAGCATTCAGACATGTATTAAGCTGATATTGAA 552

```

```

Qy      145 AlaIlePheThrArgIyGluLeuAlaGlyArgArgThrValSerPheAspSerLeuLeu 164
Db      553 GTGTGACCAATTAAGGAATATGAGGCGATCAGATCTGCACGCTTTTCAATGCTGTTA 612
Qy      165 GluGlnProGlyIyIleGluAlaAspAsnAlaPheAlaAlaThrGlyProAspThrIle 184
Db      613 GATACACCAAGTTTCA--AATGTTCAAGAGTTTATCAAAACCTTGATGAACCAAGATT 669
Qy      185 AlaIyPheLeuPheThrSerGlySerThrIyLeuProIyAlaIyProThrGln 204
Db      670 GCCAAATTTCTGTTCATCAGGTCAAGGTCAACTAAATTACTTAAGGCTGACGACACAT 729
Qy      205 ArgMetLeuCysAlaAsnGlnMetLeuLeuGlnThrPheProValPheGlyGlu 224
Db      730 TTAATGTGTGTCTTAATCAGCAAAATGTTATTCACACCTTCCCTGAGTTGAAGAAACA 789
Qy      225 ProProValLeuValAspTrpLeuProTyrrPAsnHisThrPheGlyIySerHisAsnIle 244
Db      790 CCGCTGCTCTACTCGACTGCTGCTTGGCACACACATTTGGCGGAGTCAAAATGTC 849
Qy      245 GlyIleValLeuTyrrAsnGlyIyThrTyrrTyrrLeuAspAspGlyIySerProThrAlaGln 264
Db      850 GGCATGCACTATTAACGCGGTACGATTTACATTGATGATGCAAAACCGTTGCAGCA 909
Qy      265 GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrrLeuThrVal 284
Db      910 AATTTGACGAAACATTTGCTAATCTCAAGAAATTTCTCCAACTGTTTATTTAAATGTC 969
Qy      285 ProIyGlyTyrrGluLeuValGlyAlaLeuGluArgAspSerThrLeuArgIyArg 304
Db      970 CCAAAAGTTGGGAACCTCACCGCAAGCTTAAGAAAGATGAAGAAATTAAGAGAGCGC 1029
Qy      305 PhePheAlaArgMetIyLeuPhePhePheAlaAlaIleGlyLeuSerGlnGlyIleTrp 324
Db      1030 TTTTTCGCAAGTTAAATTTATTTCTTTCGCGGTGCTCACTTTCAGAAAGCGGCTGG 1089
Qy      325 AspArgLeuAspArgValAlaGluGlnHisCysGlyIyArgIleArgMetMetAlaGly 344
Db      1090 AACAGCTGATTAATTAATTCCTACGCAACATTCGGAGAAATAATCCGCAATTATGACGGA 1149
Qy      345 LeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAla 364
Db      1150 TTGGGATGACCGAAACTGCTCCCTCTGCTTGTATTAACAATGCGCCACGCGTAATGGCT 1209
Qy      365 GlyTyrrIleGlyLeuProAlaProGlyCysGluValIyLeuValProValAspGlyIyS 384
Db      1210 GCTTATTTGTTACCTCTCTCCGGATGCGAAATTAACCTAGTTTCATGTGTGACAA 1269
Qy      385 LeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrrTrpArg--AlaProGlu 403
Db      1270 CTTGAGTTTGGCTTCGTGCGAAACGTCATGAAGGCTATTTGGCGCTTAAGCGGAGAC 1329
Qy      404 GlnAsnAlaGlnAlaPheAspGluGlyIyTyrrTyrrCysSerGlyAspAlaIleIyLeu 423
Db      1330 CAACAAAGTAACTATTTTATGATGATGAAGGCTTTTCCAAACGCGCATGCCGTTCTT 1389
Qy      424 AlaAspProAlaAspProGlnIySerGlyLeuMetPheAspGlyArgIleAlaGluAspPhe 443
Db      1390 GTGATGTCAATGATCCAAACAAAGGCTTAATGTACAGACGAGAAATTCGCAAGACTTT 1449
Qy      444 IyLeuSerSerGlyValPheValSerValIyProLeuArgThrArgAlaValLeuGln 463
Db      1450 AAATCAATACAGGCACTTTTGTGAATGTGCGCACACTACGCAACAAAGTCTTATTCAA 1509
Qy      464 GlyIySerTyrrValLeuAspValValValAlaAlaProAspArgGlyCysLeuGlyLeu 483
Db      1510 GGTAAATTTACATCAACAAAGTGTGTATTAATCATTGTTCAACCTGAATCTATGGTATT 1569
Qy      484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGly--IySerGlyAla 502
Db      1570 CTGATTTTTCAAAAATTAAGCCTTGCTGCTCAATATGACGAGCTTAAAGCTGGCGCAATAT 1629

```

Qy	503	SeraAsp1aqlu1valu1a1aSerGluProVal1ArAla1Arphea1aAsp1rleu1ys	522
Db	1630	TTGCGACACAAAGATATCGCAACCCCTTAAAGTCCAATGGTTCGCCAATTTTAAACG	1689
Qy	523	ArgLeuBna1rgrglua1Ar1rGl1yAma1a1aSerArg1lMe1rTrpAl1GlyLeuBna1p	542
Db	1690	ACCTTTAA1TAAAGATGCGATCGCAGCTCAATATACGTCTCAATGCTTTATTTAATGACC	1749
Qy	543	ThisProBose1r1eAp1y1eG1y1u1a1r1rAsp1y1eG1y1Ser1l1eAnG1n1a1rG1a	562
Db	1750	GAGCCACTCAGTATAGATGCGCGCGAAGTACCGATTAAGCACTCAATCAAGACGT	1809
Qy	563	Val1eug1n1r1rPar1grSer1a1a1y1a1rVal1a1rAla1a1eU1r1y1rG1y1G1u1a1p1G1n1Ser1n1e1r	582
Db	1810	ATTACCAAGCGTCGTGCGCTTTAATGTATGACCTTTATCAAAAGCAGACTGATTAACCG	1869
Qy	583	Leu 583	
Db	1870	CTG 1872	
RESULT 10			
ACCPAOP/c			
LOCUS	45324 bp	DNA	linear BCT 27-JUN-2003
DEFINITION	Acinetobacter sp. ADP1 dca-pca-qui-pob-hca supraoperonic gene cluster, complete sequence.		
ACCESSION	L05770	AY056039	L03407
VERSION	L05770.5	GI:32306866	
KEYWORDS			
SOURCE	Acinetobacter sp. ADP1		
ORGANISM	Acinetobacter sp. ADP1		
	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter.		
REFERENCE	1 (bases 1 to 45324)		
AUTHORS	Hartnett,C., Neidell,E.L., Ngai,K.L. and Ornston,L.N.		
TITLE	DNA sequences of genes encoding Acinetobacter calcoaceticus protocatechate 3,4-dioxygenase: evidence indicating shuffling of genes and of DNA sequences within genes during their evolutionary divergence		
JOURNAL	J. Bacteriol.	172 (2),	956-966 (1990)
MEDLINE	9013033		
PUBMED	2298704		
REFERENCE	2 (bases 1 to 45324)		
AUTHORS	DiMarco,A.A., Averhoff,B.A., Kim,E.E. and Ornston,L.N.		
TITLE	Evolutionary divergence of pobA, the structural gene encoding p-hydroxybenzoate hydroxylase in an Acinetobacter calcoaceticus strain well-suited for genetic analysis		
JOURNAL	Gene	125 (1),	25-33 (1993)
MEDLINE	93194074		
PUBMED	8449410		
REFERENCE	3 (bases 1 to 45324)		
AUTHORS	DiMarco,A.A., Averhoff,B. and Ornston,L.N.		
TITLE	Identification of the transcriptional activator pobR and characterization of its role in the expression of pobA, the structural gene for p-hydroxybenzoate hydroxylase in Acinetobacter calcoaceticus		
JOURNAL	J. Bacteriol.	175 (14),	4499-4506 (1993)
MEDLINE	93322329		
PUBMED	8331077		
REFERENCE	4 (bases 1 to 45324)		
AUTHORS	Hartnett,G.B. and Ornston,L.N.		
TITLE	Acquisition of apparent DNA slippage structures during extensive evolutionary divergence of pcad and cadC genes encoding identical catalytic activities in Acinetobacter calcoaceticus		
JOURNAL	Gene	142 (1),	23-29 (1994)
MEDLINE	94237485		
PUBMED	8181753		
REFERENCE	5 (bases 1 to 45324)		
AUTHORS	Kowalchuk,G.A., Hartnett,G.B., Benson,A., Houghton,J.E., Ngai,K.L. and Ornston,L.N.		
TITLE	Contrasting patterns of evolutionary divergence within the Acinetobacter calcoaceticus pca operon		
JOURNAL	Gene	146 (1),	23-30 (1994)
MEDLINE	94341565		
PUBMED	8063101		
REFERENCE	6 (bases 1 to 45324)		
AUTHORS	Elismore,D.A. and Ornston,L.N.		
TITLE	The pca-pob supraoperonic cluster of Acinetobacter calcoaceticus containing quia, the structural gene for quinate-shikimate dehydrogenase		
JOURNAL	J. Bacteriol.	176 (24),	7659-7666 (1994)
MEDLINE	95095936		
PUBMED	8002591		
REFERENCE	7 (bases 23320 to 27131)		
AUTHORS	Elismore,D.A. and Ornston,L.N.		
TITLE	Unusual ancestry of dehydratases associated with quinate catabolism in Acinetobacter calcoaceticus		
JOURNAL	J. Bacteriol.	177 (20),	5971-5978 (1995)
MEDLINE	96011389		
PUBMED	7592351		
REFERENCE	8 (bases 1 to 45324)		
AUTHORS	Gerischer,U., Segura,A. and Ornston,L.N.		
TITLE	PcaU, a transcriptional activator of genes for protocatechuate utilization in Acinetobacter		
JOURNAL	J. Bacteriol.	180 (6),	1512-1524 (1998)
MEDLINE	98175676		
PUBMED	9515921		
REFERENCE	9 (bases 1 to 9706)		
AUTHORS	Parke,D., Garcia,M.A. and Ornston,L.N.		
TITLE	Cloning and genetic characterization of dca genes required for beta-oxidation of straight-chain dicarboxylic acids in Acinetobacter sp. strain ADP1		
JOURNAL	Appl. Environ. Microbiol.	67 (10),	4817-4827 (2001)
MEDLINE	21455042		
PUBMED	11571189		
REFERENCE	10 (bases 1 to 45324)		
AUTHORS	Parke,D. and Ornston,L.N.		
TITLE	Hydroxycinnamate (hca) catabolic genes from Acinetobacter sp. strain ADP1 are repressed by Hcra and induced by hydroxycinnamoyl-CoA thioesters		
JOURNAL	Appl. Environ. Microbiol.	(2003)	In press
MEDLINE	11 (bases 1 to 45324)		
PUBMED	12514037		
REFERENCE	12 (bases 12470 to 15266)		
AUTHORS	Gerischer,U.C.		
TITLE	Direct Submision		
JOURNAL	Submitted (15-DEC-1993)		
MEDLINE	22401456		
PUBMED	12514037		
REFERENCE	13 (bases 27214 to 29733)		
AUTHORS	Elismore,D.A.		
TITLE	Direct Submision		
JOURNAL	Submitted (28-JUN-1994)		
MEDLINE	14 (bases 23320 to 27131)		
PUBMED	8181753		
REFERENCE	14 (bases 23320 to 27131)		
AUTHORS	Elismore,D.A.		
TITLE	Direct Submision		
JOURNAL	Submitted (25-JAN-1995)		
MEDLINE	15 (bases 1 to 45324)		
PUBMED	8181753		
REFERENCE	15 (bases 1 to 45324)		
AUTHORS	Ornston,L.N.		

AUTHORS Smith,M.A., Young,D.M. and Ornston,L.N.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2001) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
REFERENCE 18 (bases 1 to 45324)
AUTHORS Smith,M.A., Young,D.M. and Ornston,L.N.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-2002) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
REMARK Sequence update by submitter
AUTHORS 19 (bases 1 to 45324)
REMARK 19 (bases 1 to 45324)
AUTHORS Parke,D. and Ornston,L.N.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2003) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
REMARK Sequence update by submitter
COMMENT On Jun 27, 2003 this sequence version replaced gi:20492839.
FEATURES
source
1. .45324
/organism="Acinetobacter sp. ADPI"
/mol_type="genomic DNA"
/strain="ADPI"
/db_xref="taxon:62977"
/note="Formerly Acinetobacter calcoaceticus"
/complement (<1. .689)
/gene="dcaF"
/complement (<1. .689)
/gene="dcaF"
/function="putative thiolase in catabolism of dicarboxylic acids"
/codon_start=1
/transl_table=11
/product="DcaF"
/protein_id="FAL09090.1"
/db_xref="GI:15812038"
/translation="MNAIYIDGLRTRPFGHAGELASIRPDDLAGIVYIQRILKRTYA
GAIIDVIFGDTIQAGSDSRNIAIRHALLAGLVTVPGQIVNRLCASGLAIIIDSARA
ITCGEDLIYAGGVESMSRAPFVWGKAESYSDAKIYDTTIGTRPNKKIYAQYGGH
SMPTGDNVAVEGYISREQADLFPAQSQAKYQALREGFAGETITAVEVSQKKLPK
QVTEDEHPRSS"
/complement (710. .2227)
/gene="dcaH"
/complement (710. .2227)
/gene="dcaH"
/function="hydroxyacyl-CoA dehydrogenase in catabolism of dicarboxylic acids"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="DcaH"
/protein_id="FAL09091.1"
/db_xref="GI:15812039"
/translation="MTHPIKIAIIGVWGSGIAQIAASGHTIYLYAKAGAAOAA
KQAIATPQKLDKPKITTEYADPAANALLIANELDKDCDILVAILYERLDIKSL
WQLEAIVPETTILASNTSSLIATAISNCKPERVAGHFENVPDMKVENYDLSGLK
TQPKHLEINQSRVLAGHRVPAKDPGFIINAGVAGGEBALKIINENVTDISEDR
IILDGVGFRMGPELMDLIGLDVSHPMESIYHOYEEAARYNSLTQKMLEKQGR
KQQGQFYDRTGSGTSETSAKVAERLLTVPKWIAADFEEDKOLLINYLTHNIQDV
GAKPOADISICLACVGEPTTHALRLNVNPAHSVAILDMLYGEKHTTIPMSLITETLY
SHAHSIFNIDGAMVSTIGESIGFVAQRILAMVINIGCDIAQQAISVDDINAAVPLG
LGYPGPIEWGDISNKKILLINRIALTSDRYPSPWLQRRVALNLPITFTT"
gene
complement (2276. .3037)
/gene="dcaC"
complement (2276. .3037)
/gene="dcaC"
/note="putative ketoacyl-CoA reductase"
/codon_start=1
/transl_table=11
/product="DcaC"
/protein_id="FAL09092.1"
/db_xref="GI:15812040"
/translation="MARRVLIYAGSGIGRCIAEALFKQGDQVFCVDINDSHIQGRQ
DYPQLHTPCDILGQPKRYOAMFDPAVOQLGGIDLILINNTGVSGETVAABEELSFEEWTQ

VINLINSTFVTKIAIPYLKOSDAGVILNLSIAGRGYPRRLAYSTSKGIVGTK
TLMSEIGADIRVNAILPGADVGERVQRLQARADSNMISIDVTEANLANOSIKRFFV
INPKHIALDLCIFIASDSGRISGQILPIDDKORLT"
Alignment Scores:
Pred. No.: 1,31e-104 Length: 45324
Score: 1621.50 Matches: 308
Percent Similarity: 69.54% Conservative: 96
Best Local Similarity: 53.01% Mismatches: 174
Query Match: 52.61% Indels: 3
DB: 1 Gaps: 3
US-09-750-986D-30 (1-589) x ACCPCAOP (1-45324)
QY 5 GluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluIuhIeTrpAlaIeYrth 24
DB 39875 GAACAACTCAAGCCTTATCCGAAACACGACCGATCGTTGATTGATTGCTAAAGTT 39816
QY 25 ArgProGluGlnThrCysValAlaAlaArgAlaAlaAngIylGluTrpArgArgIleSer 44
DB 39815 AAGCCCTACCAATACCTTCCCTTAAACCAATACACAAAGACGAATGATCAAGCTCAGT 39756
QY 45 TyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerIleuLeuProIyrgIleu 64
DB 39755 TATGCAAAACCTTGCAACGCTGCTGGCATATTGACAAAGCTTGATACGCCAACTTA 39696
QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluIuhIeSleuGlnLeuAla 84
DB 39695 AGCCAAAGACGACCTTGGTATATGATTCGGAAAGATCTTGAAACATTAAACGCTTGG 39636
QY 85 PheGlyAlaMetCysTrpAlaGlyIleProIyCysProValSerProAlaIySerLeuLeu 104
DB 39635 ATGGGCGCATGTGGACAGTGAATTCATTTCGTAATTCGACAGCTACCTACCTGTA 39576
QY 105 SerGlnAspLeuAlaIySleuArgHisIleValIleGlyLeuGlnInProGlyLeuValIpe 124
DB 39575 TCTCAAGATTTTGGCAATTAATAACATGTGTTGATGTCTCACCCAGCATGCTGAT 39516
QY 125 AlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValPro 144
DB 39515 GCAAATGATGACACGACCTTTGGTAAAGCATTCMAAGGTGACGCAATCATATTGAA 39456
QY 145 AlaIlePheThrArgGlyLeuLeuAlaGlyArgArgThValSerPheAspSerLeuLeu 164
DB 39455 ATGTGACCAATPAAAGCATTTATAGGTGATCAAGCTGTACATCGTTTACGTTTATTA 39396
QY 165 GluGlnProGlyGlyIleGluIlaAspAsnAlaPheAlaIleThrGlyProAspThrIle 184
DB 39395 GATACCCCT--GTGACCATGTTCAAGAACATTATGAACCGTTAGATGACATCGATT 39339
QY 185 AlaIyPheLeuPheThrSerGlySerThrIySleuProIySleAlaValProThrThrgln 204
DB 39338 GCCAATTTTATTATTCATCCGGTTGACCAAGCTTCMAAGCGCTTCCGACACACAT 39279
QY 205 ArgMetLeuCyseAlaAngIlnGlnMetLeuLeuGlnThrPheProValPheGlyGluGlu 224
DB 39278 TTGATCTCTGTATCATCAACAACAACTTCAACAACTTCCAAATTTGAAGCACCA 39219
QY 225 ProProValLeuValAspTrpLeuProIyPasnHisThPheGlyGlySerHisAsnIle 244
DB 39218 CCGCTGTCTTACTGACCGTGTGCTTTCGATCATACATTCGTTGGAAGCCAAATGTT 39159
QY 245 GlyIleValLeuIyTranGlyGlyIyThrIyTrIyLeuAspAspIySleProThrAlaGln 264
DB 39158 GGGATTGGCGCTTATTAACCGCGGAACATTTCATCATGATGACGGTAAAGCCCGTCCGGGA 39099
QY 265 GlyPheAlaGluThrIleuArgAsnLeuSerGluIleSerProThrAlaIyTrIeuthrVal 284
DB 39098 AAATTGATGAACGATTCGAATTTAAAGAAATTTCCACCGCGGCTTATTCATTAATGTA 39039
QY 285 ProIyGlyIyTrGluGluIleuValGlyAlaIleuGluArgAspSerThrIleuArgIuArg 304
DB 39038 CCCAAAGGCTGGGAACCTCACCGATGCACTTGAAAGATGCGCAACTCAGAGATCGC 38979

```

QY 305 PhePheAlaArgMetLysLeuPhePhePheAlaAlaIleuSerGlnGlyIleTyr 324
Db 38978 TTCTTTCCTCAACGTTAAATTTTGTTCGTCTGCTGATTAAGCAGGCTGG 38919
QY 325 AspArgLeuAspArgValAlaGluGlnHisCysGlyGluArgIleArgMetMetAlaGly 344
Db 38918 AACCGATCGATTAATTAATTCGACAGGCATTGGCGTGAATAAATCCGATCATAGTGGC 38859
QY 345 LeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAla 364
Db 38858 TTGGGCGTGAACCGAACGCGCATCGCGCTTTACCAAGGCGCCAGCGGTAATGAGCT 38799
QY 365 GlyTyrIleGlyLeuProAlaProGlyCysGluValLysLeuValProValAspGlyLys 384
Db 38798 GGTATTATTCGCTACCTGCTCCGGGCTGTGAATCAATTAATGATTCCTATGTGACAAA 38739
QY 385 LeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTyrArg---AlaProGlu 403
Db 38738 CTGGAATTTTGTGTGCTGGGACAGCATGTAAGATTATCGCCCTTAAGCCGAC 38679
QY 404 GluHisAlaGluAlaPheAspGluGlyTyrTyrCysSerGlyAspAlaIleLysLeu 423
Db 38678 CACCAAGCACAATTTTGAAGAAGCGGCTTTCCATCTGCGCATGCGGTACGTCTG 38619
QY 424 AlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaAspPhe 443
Db 38618 GTGATCTCAACGACCTCGACAGGCGCTGATGACAGCGCGCTATTCGAGAAATTC 38559
QY 444 LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
Db 38558 AAATCTCAACGCGGTACTTTGTGAATGTGGGCACTTACGCAACAGCAGCATCTACAA 38499
QY 464 GlyGlySerTyrValLeuAspValValAlaAlaProAspArgGluCysLeuGlyLeu 483
Db 38498 GGTAAATTTGTGATTCAGAGATGTTGTATTACCGGTCAATCAATGAGCATGCGCTTC 38439
QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGly---LysGluAla 502
Db 38438 TTGGTTTTTCTTAAGCTTGAATGCCCTGTCAGATTTTCTGCTGGAATGAAAATCAT 38379
QY 503 SerAspAlaGluValLeuAlaSerGluProValArgAlaTyrPheAlaAspTyrLeuLys 522
Db 38378 AGGCGAATAGATATCTTGCACAACCCTAAGTACAGAGTGTTCACAGCTTTTAATT 38319
QY 523 ArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeuLeuAsp 542
Db 38318 CAGTACAAACAAGATGCCAGCGGAGCTTCAATCGGCTATCAATGTTATTATGACT 38259
QY 543 ThrProSerIleAspArgLysGlyGluValThrAspArgLysSerIleAsnGlnArgAla 562
Db 38258 GAAGCAACCAAGCTCGATGAGGTAAGTAACCATTAAGCCATTTAAATCAGAGTAAC 38199
QY 563 ValLeuGlnTyrArgSerAlaLysValAspAlaLeuTyrArgGlyGluAspGlnSerMet 582
Db 38198 ATCTTAAAGCTGTGTGTCAGCTGTAAGAACTGTATTAACAACAAGATTAATCCA 38139
QY 583 Leu 583
Db 38138 TTG 38136

```

RESULT 11
CR543861_17/c
WPCOMMENT
Sequence split into 36 fragments LOCUS CR543861 Accession CR543861

Fragment Name	Begin	End
CR543861_00	1	110000
CR543861_01	100001	210000
CR543861_02	200001	310000
CR543861_03	300001	410000
CR543861_04	400001	510000
CR543861_05	500001	610000
CR543861_06	600001	710000

```

CR543861_07 700001 810000
CR543861_08 800001 910000
CR543861_09 900001 1010000
CR543861_10 1000001 1110000
CR543861_11 1100001 1210000
CR543861_12 1200001 1310000
CR543861_13 1300001 1410000
CR543861_14 1400001 1510000
CR543861_15 1500001 1610000
CR543861_16 1600001 1710000
CR543861_17 1700001 1810000
CR543861_18 1800001 1910000
CR543861_19 1900001 2010000
CR543861_20 2000001 2110000
CR543861_21 2100001 2210000
CR543861_22 2200001 2310000
CR543861_23 2300001 2410000
CR543861_24 2400001 2510000
CR543861_25 2500001 2610000
CR543861_26 2600001 2710000
CR543861_27 2700001 2810000
CR543861_28 2800001 2910000
CR543861_29 2900001 3010000
CR543861_30 3000001 3110000
CR543861_31 3100001 3210000
CR543861_32 3200001 3310000
CR543861_33 3300001 3410000
CR543861_34 3400001 3510000
CR543861_35 3500001 3598621

```

Continuation (18 of 36) of CR543861 from base 1700001 (CR543861 Acinetobacter sp. ADP1 cc

Alignment Scores:

Pred. No.:	6.8e-104	Length:	110000
Score:	1618.50	Matches:	307
Percent Similarity:	69.54%	Conservative:	97
Best Local Similarity:	52.84%	Mismatches:	174
Query Match:	52.51%	Indels:	3
DB:	1	Gaps:	3

US-09-750-986D-30 (1-589) x CR543861_17 (1-110000)

```

QY 5 GluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGlnHisTyrPalaLysThr 24
Db 33726 GAACCACTCAAGGCTTATCCGCAAAAAGTACCGATGCTTGAATCATTTTGTCTAAAGTT 33667
QY 25 ArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyLutTyrArgArgIleSer 44
Db 33666 AAGCCCTAACATACCTTGGCCGCTTAAACGCAATACACAAAGCAATGATCAGGCTCAGT 33607
QY 45 TyrAlaGluMetPheHisAsnValArgAlaAlaIleAlaGlnSerLeuLeuProTyrGlyLeu 64
Db 33606 TATGACAGAAACCTTGCAACGCTGTCGATATTCACAACTTGTGATCGACGCAACTTA 33547
QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGlnHisLeuGlnLeuAla 84
Db 33546 AGCCAAAGACGACCTTGGTATTTATTCGGGAAAGATGATTGAACACTTAAGCTTTTCG 33487
QY 85 PheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeu 104
Db 33486 ATGGGCGCGATGTGGACAGGTGATACCATTTTGTGATGTCACCCGAGGATGGTAT 33427
QY 105 SerGluAspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLysLeuValPhe 124
Db 33426 TCTCAAGATTTTGGCAATTAATAAGTGTGATGTCACCCGAGGATGGTAT 33367
QY 125 AlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValPro 144
Db 33366 GCAATGATGCAAGACCTTTTGTCTAAAGCATTCAAAGCTGACGCAAGATCATATTGAA 33307
QY 145 AlaIlePheThrArgGlyLeuLeuAlaGlyArgArgThrValSerPheAspSerLeuLeu 164
Db 33306 ATTGTGACCAATTAAGCATTTATAGGTGATCAAGTCTGTACATCGTTTCAGTCTTATTA 33247

```

QY 165 GlnGlnProGlyGlyIleGlnAlaAspAsnAlaPheAlaIleThrGlyProAspThrIle 184
 Db 33246 GATACCCCT---GTGACCATGTTCAGAGAACATTATGAAACGGTAGATGACATTCATTT 33190
 QY 185 AlaLysPheLeuPheThrSerGlySerThrLysLeuProLysAlaValProThrThrGln 204
 Db 33189 GCCAAGTTCATTTATTCATCCGGTTCGACCAAGCTTCGAAAGCGTGTTCGACCAACAT 33130
 QY 205 ArgMetLeuCysAlaAsnGlnGlnMetLeuGlnThrPheProValPheGlyGln 224
 Db 33129 TTTATGCTGTGTCATCATCAACAATGCTCTTCAAACTTTCCAGATTTGGAAGACACA 33070
 QY 225 ProProValLeuValAspThrLeuProThrAsnHisThrPheGlySerHisAsnIle 244
 Db 33069 CCGCTGTCTTACTGACCTGGTGTCTTGGCATCATCATTCGATGGAACCAACATGTT 33010
 QY 245 GlyIleValLeuValThrGlnGlyValThrTyrThrLeuAspAspGlyLysProThrAlaGln 264
 Db 33009 GGGATTCGCTTATTAACCGCGGAAACATTTACATCATACGGTAAGCCCGTCCGGA 32950
 QY 265 GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrLeuThrVal 284
 Db 32949 AATTTGATGAAACGATCGAAATTTAAAGAAATTTACCGACGGTTCCTTAATGTA 32890
 QY 285 ProLysGlyTyrGluLeuValGlyAlaLeuGluArgAspSerThrLeuArgGluArg 304
 Db 32889 CCAAAAGCTGGGAGAAACTCACCGATGCACTTGAAAAAATGATCCGAACTCAGAGATCGC 32830
 QY 305 PhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSerGlnGlyIleTyr 324
 Db 32829 TTTCTTGGCCAACTTAATTTGTTTGTGCTGGTCTCATTTCTGAAAGCAGGCTGG 32770
 QY 325 AspArgLeuAspArgValAlaGlnGlnHisCysGlyGluArgIleArgMetAlaGly 344
 Db 32769 AACCGATCGATTAATTAATGACACAGCAGCATTCGCGTAATAAATCCGCATCATGAGTGC 32710
 QY 345 LeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAla 364
 Db 32709 TTGGGATGACCGAAACGACACCATCGTGGCTTTTACACAGGCCACGCGTAATGGCT 32650
 QY 365 GlyTyrIleGlyLeuProAlaProGlyCysGluValLysLeuValProValAspGlyLys 384
 Db 32649 GGTTCATTCGTCACCTGCTCGGGCTGGAATCAATTAATGATCCCTTATGCTGACAAA 32590
 QY 385 LeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTyrArg---AlaProGlu 403
 Db 32589 CTTGAATTTTGTGTCGTGCGCAAGCAGTCATGAAAGTTACGCGCCCTTAAGCGGCAC 32530
 QY 404 GlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleLysLeu 423
 Db 32529 CACCAAAAGCACATTTTGTGTAAGACGGCTTTTCCATCTGCGCATGCGTACGCTG 32470
 QY 424 AlaAspProAlaAspProGluLysGlyLeuMetPheAspGlyArgIleAlaGluAspPhe 443
 Db 32469 GTCGATCCAAACACCTGACAGGGCTGATGACACGCGGTATTCGAGAAAGATTTC 32410
 QY 444 LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGln 463
 Db 32409 AATCTACACAGCGGTACCTTTGGAATGTGGCAGCTTTACGAAACAGGACATCATTCMA 32350
 QY 464 GlyGlySerTyrValLeuAspValValAlaAlaProAspArgGluCysLeuGlyLeu 483
 Db 32349 GGTAAATTTTGTTCAGATGTTTGTGTAATTCGCGTTCAATCAATGCGATTTGGCTTC 32290
 QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGly---LysGluAla 502
 Db 32289 TTGGTTTTCCTAAGCTGATGCTGCGTGCAGATTTCGTGCTGATATGAAAAATCAT 32230
 QY 503 SerAspAlaGluValLeuLysSerGluProValArgAlaPhePheAlaAspThrLys 522
 Db 32229 AGGCGAATAGATATCTTTCGAAACACCTTAAGTACAGAGTGTTCGACAGTTTAAAT 32170
 QY 523 ArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeuLeuAsp 542

Db 32169 CAGTACAAACAAGATCCACGGCGACCTCAAAATCGGGATCAATGTTGATTTAAAGACT 32110
 QY 543 ThrProSerIleAspLysGlyValThrAspLysGlySerIleAsnArgAla 562
 Db 32109 GAGCACCAAGCTCATGCGGTGAAGTACCGATTAAGGCAATTTAAATTCAGACTAAC 32050
 QY 563 ValLeuGlnTyrArgSerAlaLysValAspAlaLeuTyrArgGlyLysArgInsMet 582
 Db 32049 ATCTTAAACGTCGTCTGACCTGCGAAGAACTGTATTAACAACAACAGATAATCCA 31990
 QY 583 Leu 583
 Db 31989 TTG 31987
 RESULT 12
 LOCUS BX572598 346879 bp DNA linear BCT 20-APR-2004
 DEFINITION Rhodopseudomonas palustris CGA009 complete genome; segment 6/16.
 ACCESSION BX572598 BX571963
 VERSION BX572598.1 GI:39648490
 KEYWORDS complete genomes.
 SOURCE Rhodopseudomonas palustris CGA009
 ORGANISM Rhodopseudomonas palustris CGA009
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Bradyrhizobiaceae; Rhodopseudomonas.
 1 (bases 1 to 346879)
 AUTHORS Lamer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L.,
 Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S., Tabita, F.R.,
 Gibson, J.L., Hanson, T.E., Bobec, C., Torrez, J.L., Perez, C.,
 Harrison, F.H., Gibson, J. and Harwood, C.S.
 TITLE Complete genome sequence of the metabolically versatile
 photosynthetic bacterium Rhodopseudomonas palustris
 Nat. Biotechnol. 22 (1), 55-61 (2004)
 JOURNAL 14704707
 PUBMED 2 (bases 1 to 346879)
 REFERENCE Lamer, F.W. and Harwood, C.S.
 AUTHORS Rhodopseudomonas genome consortium
 CONSRM Direct Submission
 TITLE Submitted (24-JUL-2003) Submitted on behalf of the Rhodopseudomonas
 genome consortium, the DOE Joint Genome Institute, Production
 Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
 USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
 1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
 lamerf@ornl.gov
 FEATURES
 source Location/Qualifiers
 1..346879
 /organism="Rhodopseudomonas palustris CGA009"
 /mol_type="genomic DNA"
 /strain="CGA009"
 /db_xref="taxon:258594"
 127..1005
 /gene="dapA"
 /locus_tag="RPA1571"
 127..1005
 /locus_tag="RPA1571"
 127..1005
 /gene="dapA"
 /locus_tag="RPA1571"
 /EC_number="4.2.1.52"
 /function="InterPro IPR002220 COGS COG3229"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="putative dihydrotolipicolinate synthetase"
 /protein_id="GI:39648491"
 /db_xref="GI:39648491"
 /translation="MMADPHGVPPYLVSPVDAEGRRADVMGRLCDLIQAGVHGTP
 LSTSEFALVGAQREAVRATTEAQRVPPVAVGASTSVADVAQALYELGADG
 ILALIEAYPLDQAQIESYFRAIDAEVPIVYINPQPSDLTLIDVLAERPI
 RYIKASTNTGLSLINCGDALGVFSASAIIPAAVMIGGVAGGACIARQSV
 ALYEICKAQRMEALMLQKRWVNEAFKFMILACIKKGLALQGVDDVPIPPQAL
 TAEERKAVYKVLAEIA"
 1230..3560
 /locus_tag="RPA1572"
 gene

CDS
1230..3560
/locus_tag="RPA1572"
/function="InterPro IPR000674 COG6 COG1529"
/note="Observed by proteomics
Citation: Proteomics from Verberkmoes et al. (2003)
unpublished"
/evidence=not experimental
/codon_start=1
/transl_table=11
/product="putative carbon-monoxide dehydrogenase large
subunit"
/protein_id="CAE27013.1"
/db_xref="GI:39648492"
/translacion="MNILPGSMRFAGQPVKRLDORLLTGHGGYLDLDDKAAEGALMLV
VMSPHAHADIVSIDTEAAKGMPEVAVLIGADLIADDVGTITPLPIFKRPGDSPLML
PLRLLAHEKRVFGEVPAVAVATSQLAAMAEATVVEHELPATVDPRAALPOGAP
AVAEFTDNTVAAMSTGDAAKVDAPASAAHTVSLVDSORLIPSAEPSTVLAIEK
KTRGLIHTQSQTPTGTRDALEAILKRPRESIQVLVADIGGGFGQKTVYPEDALVA
YAAVKLNKRIKRMGERTERDEFIGTHGRDLSTASFLADSKGKVLAVRTSIGTGAVL
AGAGVILPLVGFVOTGVVHLPVVHFDIKAMVHTPAVGAYGAGREAVYIYERLM
DKAARKLMDPRAIRKINYIKPSQLPYTNVGVGVDSGAFAHMMDRRAELSDMDGFPA
RKAARKKGLVGRGVTSYETGGRHTENYKHAIEGRVLIHSTQAMGCGLETT
YTMIAALEIPLDKIDVGSNTDLAIGFSVGSRLVFGTAVAVSSABMIAAREK
AANILEASIDIEVSGMLTIAGTRKISLFEIAAEKGRSLVSDTGEVDGSPWPG
AHICEVDEBETGVSRVRYTTVDVGNANVPMIVAGQIHGVAGVGAALYEAAN
DDGQLTASYLYDCIPRADNLPIVNTLPSAPCKTPIFGAKGCGESGALGPPCVVH
GVLDALALPGVTLTNPITREKVMRAIQDAKSAQA"
3638..4252
/locus_tag="RPA1573"
3638..4252
/locus_tag="RPA1573"
/function="COG6 COG1704"
/note="Observed by proteomics
Citation: Proteomics from Verberkmoes et al. (2003)
unpublished"
/evidence=not experimental
/codon_start=1
/transl_table=11
/product="leuam family"
/protein_id="CAE27014.1"
/db_xref="GI:39648493"
/translacion="MRKFTLVLAALALSLNSGNYTIOSEDEYKSTWSEVNOYOR
RADLVNPLVNSGFAQOEKDLVIGTNARAKGSIQATPEVINDPAPAFQOAOGE
LSSALSRLLIVTENYPOLKSDLEFKNMAOLEGNNITVARNRYIKAVAAVNTVRS
VPTFTAMGCGYKERNFTYENKAISTAKVDTPAPAPAK"
3638..3700
/locus_tag="RPA1573"
/note="Signal predicted by SignalP 2.0 HMM (Signal
peptideprobability 0.998) with cleavage site probability
0.396 at residue 21"
4257..5180
/locus_tag="RPA1574"
4257..5180
/locus_tag="RPA1574"
/function="COG6 COG1512"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="DUF477"
/protein_id="CAE27015.1"
/db_xref="GI:39648494"
/translacion="MTAATRAVPAAGLALLCVNALALAPPAADVAVVTLGRVVD
QRTSLSDAVARLDOKLAKFAEARKGSIATVLTPTTPEAIEQFSIRVAEMKIRKX
IDGAILLVAKNDRKRIEIVYGEGALPVTAKRIIDETITPEKTKGDDGIGEVAG
DMRIGVDEGPLTPPOQHMAAPSEMDDELTGMAIPLLFGLTVNGFLKALGRVAS
GLGIIVGAVAMLFGVWYIALVAGFAFLALVITDLFGPTISSGRGSGGWTWGG
SSGSSWSGSSDGSFGSGGSGFGGASGSM"
4257..4352
/locus_tag="RPA1574"
/note="Signal predicted by SignalP 2.0 HMM (Signal
peptideprobability 1.000) with cleavage site probability
0.947 at residue 32"
order(4290..4358,4824..4880,4899..4952,4962..5030)
misc_feature

gene
/locus_tag="RPA1574"
/note="4 probable transmembrane helices predicted
byTMHMM2.0"
5194..5691
/locus_tag="RPA1575"
5194..5691
/locus_tag="RPA1575"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE27016.1"
/db_xref="GI:39648495"
/translacion="MGIIRIGKGLFTSRGAVHRAFSPDAFDAIERAIKASETHVGOI
RFVVEGLDGAFLFRDQVPERALDI FSOLRIWDETHNGVLVLLADRNVEIADR
GIDRGACADTWEAICHEMEFEREGFEGVLTGIDIRITHLAHEHPRGSPSNEIPD
APVVV"
complement(5773..6384)
/locus_tag="RPA1576"
complement(5773..6384)
/locus_tag="RPA1576"
/function="InterPro IPR004045:IPR004046 COG6 COG0625"
/note="Observed by proteomics
Citation: Proteomics from Verberkmoes et al. (2003)
unpublished"
/evidence=not experimental
/codon_start=1
/transl_table=11
/product="putative glutathione S-transferase"
/protein_id="CAE27017.1"
/db_xref="GI:39648496"
/translacion="MKLYDSIGPNPRLVRMFIABKGLTIPTOTLDIRAGERRKREHLA
RNPHGOTPTLELNGSFLAEVTAICVYIEKHQOPPLIGTPEERAEBCRWTRRIDLY
ICEPLTNGRRFPAELGKMFQDRVYCVBBAELGLKIAADRIRMMAEQMKGEYVCGARL
TLADLLYGLWLDGAQVGOPLAPANTTIAEFKRIARPSAKA"
6593..7564
/locus_tag="RPA1577"
6593..7564
/locus_tag="RPA1577"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE27018.1"
/db_xref="GI:39648497"
/translacion="WDBHREPTGEOQTGAATAATDGAONAGTPEASSASAPQSTPAN
TPNAGAAKAPDEAKPAGKSEIEKSDVSRPRPOTVTIMASPREDLYRHWDDIEVVA
PDEASQSAVHRFCVRKLAIVAAVILAIAGAGALATSLSHRSDAEFTTAASQVYK
TLDAQVLAQOIELSALAKADLDHAAKOSTAOLGTATRIEKLERAQAEPAKIAKLSST
VEKLTFRNOQPPAPAPAVNVAIVPADVGSITPPKBAKRGRLPTVEGVLDVYDG
GAVBGRSGTVEVYAADPPIRGVGRVDAIRQDGRWVYVTSRGLIVAR"
6947..7015
/locus_tag="RPA1577"
/note="1 probable transmembrane helix predicted
byTMHMM2.0"
7593..7621
/note="Transferm prediction, 98+%"
complement(7721..8494)
/gene="fpr"
/locus_tag="RPA1578"
complement(7721..8494)
/gene="fpr"
/locus_tag="RPA1578"
/EC_number="1.18.1.2"
/function="InterPro IPR001433:IPR001834 COG6 COG1018"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="ferridoxin--NADP+ reductase"
/protein_id="CAE27019.1"
/db_xref="GI:39648498"
Alignment Scores:

PUBMED
REFERENCE
AUTHORS
2
12597275
Kaneko, T., Nakamura, Y., Sato, S., Miamisawa, K., Uchiumi, T.,
Sasamoto, S., Matanabe, A., Idegawa, K., Iriuchih, M., Kaweshima, K.,
Kohara, M., Matsumoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M.
and Tabata, S.
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)
DNA Res. 9 (6), 225-256 (2002)
JOURNAL
MEDLINE
22485002
12597279
3 (bases 1 to 299950)
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp,
URL: http://www.kazusa.or.jp/~chikobae/
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
Location/Qualifiers
1. .299950
/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
36. .278
/gene="bcr7655"
36. .278
/gene="bcr7655"
/note="ORF ID: bcr7655"
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAC52920.1"
/db_xref="GI:27355939"
/translation="MMVQATDVSDSSAIDEMVAAYVDRFGRLDVVNNAGVHGGAPV
SITDKKRWTVMSDVGVFGCAALTLRPGARS"
275. .637
/gene="blr7656"
275. .637
/gene="blr7656"
/note="ORF ID: blr7656"
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAC52921.1"
/db_xref="GI:27355940"
/translation="MNRVSVSGTGDGMSPPYNAKGAVVNLTPALALDGGKIRVN
AVCPSLTRNGITEDMDKELAKFAERITLGRVCEBEVAAVIAFLASDASMTGA
NNAVDSVSNAGPOFE"
complement (660. .1346)
/gene="bll7657"
complement (660. .1346)
/gene="bll7657"
/note="ORF ID: bll7657"
putative phosphoglycolate phosphatase"
/codon_start=1
/transl_table=1
/protein_id="BAC52922.1"
/db_xref="GI:27355941"
/translation="MMDESELPKAIPLDGLTLDSDVHLALAOEAMVERGHDT
FEVRSQIGKGLTFLVLSDEQREHGKLEWRGNRRKSTYLPILRPSASQL
QVRVREGLRIAVASAKDELKTYLAIASDLVDVTSSEVDSEKRPADIEVVLK
KLGIKSGEVAIVGTPYDAIAARAGVSAIGVLCGFTENSLREAGCVEVYPGPTLY
ARFGDTALTK"
complement (1503. .1880)
/gene="bll7658"
complement (1503. .1880)
/gene="bll7658"
/note="ORF ID: bll7658"
/codon_start=1
/transl_table=1

gene
CDS
/product="two-component response regulator"
/protein_id="BAC52923.1"
/db_xref="GI:27355942"
/translation="MSNALRYLVBEALICSPFEDALSDGGEACSVHSGEAMSTP
RDERECRALLTPVNGDGISGWEALRQIREITPGFPVYMTSASAPDMKSGQVDSL
LIRKPPAQQLAAVSQLDTRA"
complement (1987. .2319)
/gene="bll7659"
complement (1987. .2319)
/gene="bll7659"
/note="ORF ID: bll7659"
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAC52924.1"
/db_xref="GI:27355943"
/translation="MMEDRETSLSGSDNVECTNYGADGEKGYTERIMDKVSGK
VSHAIVSPDGLRIGVHYPLPWAQLKYDTNLGVTGILTDLRSPAKYADESSMMW
SDPAPRAP"
complement (2397. .2825)
/gene="bll7660"
complement (2397. .2825)
/gene="bll7660"
/note="ORF ID: bll7660"
unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAC52925.1"
/db_xref="GI:27355944"
/translation="MRRPRMAKFCRIILDTTIVPNNLRLOGAAAAFRCIAGTLDLGR
ACRQSVAAEVRISIVQGTPEPISRLRGRRSPRVSDPVFDIAGMKRFQFHTELHQ
PGHDLVHERRAPGFAEOLPOKFAQLQGFHARTPLIQN"
complement (2979. .3422)
/gene="bll7661"
complement (2979. .3422)
/gene="bll7661"
/note="ORF ID: bll7661"
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAC52926.1"
/db_xref="GI:27355945"
/translation="MLMGEEFALNYFLDFGKRVSLRPMRWDFEYDESGRPVSL
VRLLIRLRORDHIRQGTVEFFDFEEDMYOQFGILLFARHDGAQYTLIATNFGDTRIV
PFWPRKAGNYVERLHGVALDLKNVPAALQVWLDIPSHYGVMLT"
3427. .3654
/gene="bcr7662"
3427. .3654
/gene="bcr7662"
/note="ORF ID: bcr7662"
unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAC52927.1"
/db_xref="GI:27355946"
/translation="MPLBARSGISAKLSRMLHGRVYGGGRISVHGKAVHYKANG
SHGRYPPDRARGSARFRSSGSSPGLQRM"
complement (4162. .4824)
/gene="bll7663"
complement (4162. .4824)
/gene="bll7663"
/note="ORF ID: bll7663"
unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAC52928.1"
/db_xref="GI:27355947"
/translation="MKRWRYASTSTVIAGVLCAGLVVGGSHVPSDAIASSVTRAP
ELMERAMHPVAATFORHDMQNSGRCGPAAYANAYRSLGEAARTGKVLACTGSCW
TGVCLMGTLDELAKVAQITVSRKTVLRULDSNOFLBHRANDPGRKITVFPDRK
IFGAGSGHSPIGGYFEABDLVFLVDFNFGPWLVERKRLFDAVNTLDDGKRGILL
IE"
complement (7319. .9061)

gene

CDS
 /gene="b117664"
 complement (7319, .9061)
 /gene="b117664"
 /note="ORF ID: b117664"
 putative adenylate cyclase 3 (EC 4.6.1.1)"
 /codon_start=1
 /transl_table=11
 /protein_id="BAC52929.1"
 /db_xref="GI:27355948"
 /translation="MTEQRYQRLAIIADVGVYSALIMORAEATYAFERLRELI
 EPSLSHDEGLIKITGDDGALAFRSPILAAFCATMEMDHAEGSGPKRIAGINLGV
 IVQDDGLVDGDNINAVRLFGIADPGGILMSERVISEBKGDLVGFEDRGEQJKNIA
 KPVRAVAVRAGASSALTERLSAARPLPAKSLIAPLPENNSGDPDEQEFAYLVEEIT
 TALSRKMLFVIAKNSFTYKGAIVNVKQVHELVYVYILGSGVRKAAGVYRITAGQI
 DATGTHIMADREEDLTDFALODVVALVDSVQPMQTEITATARSPENITAY
 DFLRSKMOQYVLTFRSLAEATKLARALIEDRFSAAALADCHASNVALYANNP
 EPPAKAIRLARIALSLDDGDPDPLSRVANVSAQVVDSPAEIEMADRAVALDPSMR
 TNNRGMVYTMAGLPEBAIRSFERRAIRMSFLDPLIMTFIGIHALVGLRRPDEALIV
 LKRALRQNPSPMPACRGLASAFALGRLDADAHAARVLEIDPTFTISGIIARVPKSA
 KMFIEGLRKAKLPE"
 9440..10444
 /gene="b1r7665"
 9440..10444
 /gene="b1r7665"
 /note="ORF ID: b1r7665"
 putative methyltransferase (EC 2.1.1.-)"
 /codon_start=1
 /transl_table=11
 /protein_id="BAC52930.1"
 /db_xref="GI:27355949"
 /translation="MTSNPWLTLGLINGFOVQAIVHASTLIADHLKQDSASBEL
 APLTQSPGALVYRLALAAVGVFEEDERFALTMDGLRTDSDATPICAMAEVCS
 PYWSTMGHLLIRTGEMAFOSLNGKDIWNYRAERPEEASVDFDRAMTESRGAELV
 IGAVDSAPHHVVVDVGGSGRLMAAILTAPRMOGTLFDQPGVIAGBAVALFAHGVD
 RCMVSGSPFESVPEBGDAIVMKVIVHMDDDALIALKRCRAMETAKULIERII

Alignment Scores:

Pred. No.: 8, 6e-88 Length: 299950
 Score: 1401.00 Matches: 293
 Percent Similarity: 64.53% Conservative: 80
 Best Local Similarity: 50.69% Mismatches: 199
 Query Match: 45.46% Indels: 6
 DB: 1 Gaps: 4

US-09-750-986D-30 (1-589) x AP005963 (1-299950)

QY 1 MetArgSerLeuGluAlaLeuLeuProGluArgGlyLeuGluAlaGluGluHis 20
 Db 199280 CTGCGGCGGAAGCAGCGCGCTCGGAGCTATTCGTCGCGATCACCGACCGTTCGATCAT 199339
 QY 21 TTPAlaLeuThrArgProGluGluGlnThrCysValAlaAlaAlaArgAlaAlaAenGlyLysTrp 40
 Db 199340 TGGGCGACGACGACGCGGACCGCGCTTCATGCGGAGCGGAGCGCGCGCGCTGG 199399
 QY 41 ArgArgGlySerGlyAlaGluMetPheHisAsnValArgAlaAlaLeaGlnSerLeuLeu 60
 Db 199400 CGCAAGATCACTATGCGCGAGCTGCTCACCGGACCGGCGACATCGCGTGGCGCTGATC 199459
 QY 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuLeuValSerGlyAsnAspLeuGluHis 80
 Db 199460 CACCGCGGCTTTCGCGGCGACCGCGCGGTGTCATCTCTCCGCAATTGCAATGCACAT 199519
 QY 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyLeuProTyrCysProValSerProAla 100
 Db 199520 GCGCTTTGGCTTCGCGCGGCTTCATGCGGCGCTTCGCTTCGCGCGGAGTGGCGGCG 199579
 QY 101 TyrSerLeuLeuSerGlnAspLeuAlaLeuAlaArgHisIleValGlyLeuLeuGlnPro 120
 Db 199580 TATTGCTGAGTGTGCAAGATTCGCGACATCTCCATCTGATGAAAGCTCTTGACGCC 199639
 QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIle-----GluThrIle 138
 Db 199640 GGCCTGCTTCGCGCGGACCGCGACAGTTCCTCGAGTGGCTCGCTGCCAATGTTTCG 199699

QY 139 LeuProAspAspValProAlaIlePheThrArgGlyLeuAlaGlyArgArgThrVal 158
 Db 199700 CTCGGACCGAGATTCGCGCTCTAC-----GGCAATGTGGCGGCGGACGTACG 199753
 QY 159 SerPheAspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAla 178
 Db 199754 CTCCTGCGGACCTTCGAGGAGCGGAGTCCGGGAGCTCTCCACAAAGTCCACGGCAG 199813
 QY 179 ThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuProLys 198
 Db 199814 ATGCGCCCGATTCGATTCGGAATTCCTGCTGACCTCGGCTCGACCGGCAATCCCAAG 199873
 QY 199 AlaValProThrThrGlnArgMetLeuCysAlaAsnGlnGlnMetLeuLeuGlnThrPhe 218
 Db 199874 GCGGTATCAACACCCAGCGCATGATCTGGCCAACAGGTGATGCTGGCCGAGACCTTC 199933
 QY 219 ProValPheGlyGluGluProProValLeuValAspTrpLeuProTrpAsnHisIleThrPhe 238
 Db 199934 GCCTTCCTCAAGGACGAGCGCGCGTCATCTGACTGCTGCTGCGAATCACACCTTT 199993
 QY 239 GlyGlySerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspArg 258
 Db 199994 GCGCGACACCAATTCGAGCTGACCGCTTCACATGGCGAGCTTCGATGTCGACGCC 200053
 QY 259 GlyLysProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerPro 278
 Db 200054 GCGACCGCATGCGCGGTGGCATCGAGAGACCGTGGCATTCACAGAGATTCGCGG 200113
 QY 279 ThrAlaTyrLeuThrValProLysGlyTyrGlnGluLeuValGlyAlaLeuGluAlaGlyArg 298
 Db 200114 ACCGTATTTTCAACCTGGCGAAGGGCTAGATGCTGCTGCTGCTTATTCGCGGACGAC 200173
 QY 299 SerThrLeuArgGlyArgPhePheAlaArgMetLysLeuPhePheAlaAlaAlaGly 318
 Db 200174 CAGGCTCGGCGCGGAATTCCTGACCGGCTGCATGCGATGTTCTTCGCGCGCGG 200233
 QY 319 LeuSerGlnGlyIleTTPAspArgLeuAspArgValAlaGluGlnHisCysGlyGlyArg 338
 Db 200234 CTGTCCCGCTGCTGTCGGGACGCTCGACGATGCGCGGTGAAGAAAAGCTATCGC 200293
 QY 339 IleArgMetMetAlaGlyLeuGlyMetThrGlyThrAlaProSerCysThrPheThrThr 358
 Db 200294 GTGCGCATGTGACCGCGCTGGCGGACCGGACCGCGCTTCTTCATGTCGCTCAAT 200353
 QY 359 GlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGlyValLysLeu 378
 Db 200354 CCGCGACACGCGCGCTCGGACATGTGCGCTCCGCTCCGGAACGACGCGCAAGCTG 200413
 QY 379 ValProValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyr 398
 Db 200414 GTCCCGAACACGCAAGCTCGAGGTCCGCGCAAGGGCGCAACGTTGATGCCGCTAC 200473
 QY 399 TTPArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGly 418
 Db 200474 TGGCGCGAGCGGATTCACCGGAAATCTTCGAGGAAGAGATTTCACAAATTGGGC 200533
 QY 419 AspAlaIleLysLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspAlaArg 438
 Db 200534 GACGCGCTCAAGGCTCCGATCCGATGATCTCACCGCGGCTTCGATTCGACGGCGC 200593
 QY 439 IleAlaGluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuAlaGlyThr 458
 Db 200594 GTACGAGAGACTTCACGCTCGGACCGGATCCTGGGTCAAGTCTCGGCCCTTGGCGCGG 200653
 QY 459 ArgAlaValLeuGluGlyLysSerTyrValLeuAspValValAlaAlaAspArg 478
 Db 200654 CGCTGTGTCGCGCTCGGACACGCTGATGCGGAGAGTGTGATCCGCGGCATCAACCGC 200713
 QY 479 GluCysLeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeu 498
 Db 200714 GACGAGGTCTCCGCGTGTGCTGCTGACCTTGACGCGGTGATGCTGTGCAACCGGACG 200773

Oy	499	Glytysglaiaaaseraplaagluvalleuualasertupro---ValArAlaTrphe	517
Db	200774	---CTGCCGGCGACGACCTGCTGTCGCGACCGCGCGCTGCTGCGAGCCCTTT	200830
Oy	518	AlahepTrleuYsArGleuansargluAlaThrGlyAsnAlaaseArGlleTrrp	537
Db	200811	CGCGAGCCCTGACCCGTTTCTCGGCTCGGCGACCGGTTCTTCGACCCGATCAGCGGC	200890
Oy	538	ValGlyleuLeuAepThrProProSerIleAplyGlyGluValThraPlyeGlySer	557
Db	200891	GCATCTCCGATGAGACCCCGCTCTCGATCGACAAGGCGAGGTCAACGACAGGCGCTCG	200950
Oy	558	IleasnglaaGAlaValleuGlnTrpArgSerAlaYsValAspAlaLeuTyr	575
Db	200951	ATCAACGAGCGCGCTGTCGAGCATCGCGCGCGCTGATCGACGACTCTAC	201004
RESULT 14			
LOCUS	AE008067/c	10173 bp	DNA linear BCT 28-MAY-2004
DEFINITION	Agrobacterium tumefaciens str. C58 circular chromosome, section 125 of 254 of the complete sequence.		
ACCESSION	AE008067	AE007869	
VERSION	AE008067.1	GI:15156484	
KEYWORDS			
SOURCE	Agrobacterium tumefaciens str. C58		
ORGANISM	Agrobacterium tumefaciens str. C58		
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.		
AUTHORS	1 (bases 1 to 10173)		
TITLE	Hinkle,G., Slater,S.C. and Goodner,B.		
	Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 10173)		
AUTHORS	Hinkle,G., Slater,S.C. and Goodner,B.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA		
FEATURES	Location/Qualifiers		
source	1..10173		
	/organism="Agrobacterium tumefaciens str. C58"		
	/mol_type="Genomic DNA"		
	/strain="C58"		
	/isolate="Cereon"		
	/db_xref="taxon:176239"		
gene	complement(100..1263)		
	/gene="AGR_C_2609"		
CDS	complement(100..1263)		
	/gene="AGR_C_2609"		
	/note="branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein - Deinococcus radiodurans (strain R1)"		
	/codon_start=1		
	/transl_table=11		
	/product="AGR_C_2609p"		
	/protein_id="AAK87205.1"		
	/db_xref="GI:15156485"		
	/translation="MGKTKMKRINTMTATVALTTTGAAPADVTIGTIPPTGPGA		
	ALGIPLKQISIFWPAIDIGEKTLKILLDDADGPPSVATTNNARFASDDKVDVIVSALT		
	PAALAVAGVAVETEPHFPALSPLEPTAKKNFVMPQSTVLVSMFARMOKDGVKRV		
	AFIGADSYGOOWDALKATKDYGLEVAERARADTSVSGVLSLARPAIFI		
	GASGTGALPOIGVERGFKGVYVSGATVDFELRIGKAVDNTVFSGVMAEDQ		
	PDSAPTKOEGLAIVKAPGKGPNTRTQFGSHWDMVWILKKAVALKKAKPGNAEF		
	RAALRDAETSGPNATHGVEYEFADNHYGIDQGVVLLTAADKFLVLPNK"		
	complement(1317..2960)		
gene	/gene="AGR_C_2611"		
	/complement(1317..2960)		
CDS	/gene="AGR_C_2611"		
	/note="probable acyl-CoA dehydrogenase - Deinococcus radiodurans (strain R1)"		
	/codon_start=1		
	/transl_table=11		
	/product="AGR_C_2611p"		
	/protein_id="AAK87206.1"		
	/db_xref="GI:15156486"		
	/translation="MSSSGMMWARSACSGRDVMSFTQATMTLKLATSGMRRIQSLR		
	GDGDDALVVALILNEARFAETVLEPLMTPADRGCTVADRVVPTGFADAFRRHAD		
	GMLAMDPEERGGQALPLVQAAQAPLFEBCVLAAMSGSTRAACHLLETREGAVA		
	DEWVPRLAAGEMAATTCISRPDAGSDALRRSVAPEGMARIEGRKTWISFGDHMA		
	MRICHLIARTGETGTGSGSLFVPEIDGENGVYVERIEEKMGHGSPCLARPD		
	GANGLLDEBERGLAQIPAMIELRLDTGCGGGLASAIIDIBETIAQERQGGDPK		
	PAYVAIHHPDIRQOLQEARLRTTLRAATLELATVMDLAMEKEBELGDLALRWL		
	LPVKNKGADAGFTVSNQAOIQIGAGAYREWEPLQYLRDARVWTIYEGTGQAOIDL		
	LTRLMEEGKGLDVFLFKRARELADYRSQITAAASVLDFAFESRRMAALENEPET		
	GLMNANNVLAAMAFAMMTVAVEGADAGARTGAA"		
gene	complement(2914..3687)		
	/gene="AGR_C_2613"		
CDS	complement(2914..3687)		
	/gene="AGR_C_2613"		
	/note="probable short-chain dehydrogenase PA2554 (imported) - Pseudomonas aeruginosa (strain PAO1)"		
	/codon_start=1		
	/transl_table=11		
	/product="AGR_C_2613p"		
	/protein_id="AAK87207.1"		
	/db_xref="GI:15156487"		
	/translation="MNTEGALVTGSAAGLGAANVLAARGAVTIPDNGBAGKK		
	LAEBIGKAVQGVDTSDADQAAIKVAASKGGLRIIVNCAGIGTAGRIILGRSGPOL		
	GDEEQVIRVNLIGTFNNMRLAAMHMEREDQSDKQDNGVIVNTASVAFFGQIQ		
	AAVAASKGQIVSLALPARABLARFRIRVNTVAIGIPLTLOGLPEVDSLAGQIPH		
	PSRLGDPAPFADVTVFRLENDYNNGEVIRUDGAIIRMQPR"		
	complement(3684..5459)		
gene	/gene="AGR_C_2614"		
CDS	complement(3684..5459)		
	/gene="AGR_C_2614"		
	/note="(AJ238746) feruloyl-CoA synthetase"		
	/codon_start=1		
	/transl_table=11		
	/product="AGR_C_2614p"		
	/protein_id="AAK87208.1"		
	/db_xref="GI:15156488"		
	/translation="MIRWNPDLGPYPQKLNERRLLHMCSPARTWADRGREPRRV		
	SYAELDKIRIGQFLDHDLSVERPLLVSENSIEHALVLAOHVGIASAITPAY		
	ATGADLTKLADIRGOITPRGVPAPEDATPPRAIGEVDDCTPTVGLNLPEDSNTPH		
	PEITLLETPEDANDRADAVGPTTVAKFLFTSGTGSFKAVITQORMLCSNGEMLAD		
	YGFFREPPVAVVMAPNWHTAAGNKVNLVYNGGYTYIDRGKSPAQIQTDNLND		
	ISPTWYFNVPAGHEMLVQAMRKDEALCRSFPRDLKMLYMGAGCAQHTALTELSPA		
	TVGHAIVLMGALGSTETAP		

AAEENRARAVIDFGHDHFCAGLDLGEHIIHKTPIEGVGRSRRWHAHVSOIENGTITPV
SLAHGVVGGGLEIASTHLI FVADBSAPALPEGCGI FVGGGSVRIARILGVAMS
DMLTGRVVASAEERWNLVOYVPRGALAKOELATTI AASNAVPI SNTAVINALPRI
QMAKEDGLFVSPSIFSSPTAASPEAEERLRAFLERKATILUKADTTA"

gene
CDS
complement (6511..7506)
/gene="AGR_C_2619"
complement (6511..7506)
/gene="AGR_C_2619"
/note="(A01317) Metf protein"

/codon_start=1
/transl_table=11
/product="AGR_C_2619p"
/protein_id="AAK87211.1"
/db_xref="GI:15156491"
/translation="MYERKILIEVAVSCILHNOVSLREKXAMGMSFMDNSGADGDVYS
SNLAWGSTEWMPEPTAKIDDPNLLPDCRRVYIAIDGPRIADWSTKRLHDEGP
VPHPLDARGISGKNELEMDKRYOGAQTQALLCGGQKOPAGSSSSIDMETETLF
DKWGFTHLHAGHPEGNRDIDSDSSRGVQALLMODPASRTDARLALTTOFAFEAK
VVIEMAKRIGNMGVITIPVHVGLAGPTQLTKFALACGVPTLKVLOKRALDSLRF
VVEPFDFAELIDHKAANDSLIEHVHIFPLGGIKRPSAMMKKHAISTGLANVA"
7739..8473
/gene="AGR_C_2620"
7739..8473
/gene="AGR_C_2620"
/note="probable transcription regulator PA4906 (imported)
- Pseudomonas aeruginosa (strain PAO1)"

gene
CDS
/codon_start=1
/transl_table=11
/product="AGR_C_2620p"
/protein_id="AAK87212.1"
/db_xref="GI:15156492"

/translation="NRQVDAATHGRAVIELREKILSGELPGSMRLEFVSTAEILDIS
RTPVRLRLTEBGLINLRPGGFYVRBPGADVDAIEVGMSTARLAEBGP
SKYALEIATYQQLDCGDRVDYDPFGYALNIPHHQALCGSEMIRREYBRA
SSLPFASPAFLPDKANIGAFRSLRGAEOHKAIVAAIVAREGALEAVAREHSTRA
RTMLEYMRBAPLIAQVPEGLAISD"
8623..10044
/gene="AGR_C_2622"
8623..10044
/gene="AGR_C_2622"
/note="PROBABLE AMINOMETHYLTRANSFERASE (GLYCINE CLEAVAGE
SYSTEM T PROTEIN)"
/codon_start=1
/transl_table=11
/product="AGR_C_2622p"
/protein_id="AAK87213.1"
/db_xref="GI:15156493"

gene
CDS
/transl_table=11
/product="AGR_C_2622p"
/protein_id="AAK87213.1"
/db_xref="GI:15156493"

/translation="MTMANSSTLEVLKQYPNVEMLRNSKIGMYVYPVAVGEFSNRS
EQAAWEGAVLPDOSHMDLIVEGDAEKFLAHGINSFNSDILNRAKHVYVTPDG
HVIYDHIIFRERDKEFLVGRAPTSMLMFCAYKMYRIKIDPRSPPRBERVLR
THYRVOIQDPDAKPIFEKNKGPIPIKPFHVDMINIGSKVQALSHGMSGAPLEW
GPEYEDSYILSCLEAKAQGVNLVCGSRAYSTNTLBSGMIPSPILGITYGGMKID
YRDLGADSYEATGALIGSFVSENIEDYVNPPELDYFTIGMKODFIKSKALEMK
GKTHRRKVTPEMNAEDVDVIAAFRPGEDHYKMDIPQANVASTADILKGGKIVG
MSFTGYSYNERCILSIGIVNDIQIGDVLITVWGBPDGSAKSTERHKQKXIRVAV
SERPYAAERADNVADSWRTKAS"

ORIGIN

Alignment Scores:

Pred. No.: 4.56e-82 Length: 10173
Score: 1294.00 Matches: 266
Percent Similarity: 61.19% Conservative: 84
Best Local Similarity: 46.10% Mismatches: 220
Query Match: 41.99% Indels: 2
DB: 1 Gaps: 2

US-09-750-986D-30 (1-589) x AE008067 (1-10173)

QY 5 GIUALLeuLeuProPheProGlyIleuLeuGluArgLeuGluHisTTPAlaIySthr 24
Db 5444 GATCCGCTGGCCCTTATCGGAGAAATGAAATGAGGCTCTTCACTGTGTCGTGG 5385
QY 25 ArgProGluGlnThrCysValAlaIaArgAlaIaIaasnGlyGluTTPArgArgIleSer 44

Db 5384 GCACCCGAGCCGACATGATGCGCCGACCTTCAGGGTCGCGAGCCATGCGCGCGGAC 5325
QY 45 TTTAAGluMetPheHisasnValArgAlaIaIaGlnSerLeuLeuProTyrGlyLeu 64
Db 5324 TACGCCGAGCCTCTCGACAAATACGCGCTATCGCCCATCTTCTTCCATCATGACCTT 5265
QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeuAla 84
Db 5264 TCGGTGAGAGCCGCTGCTGATGCTTCTGAATAATCCATGAGCATCGTTGATGCTG 5205
QY 85 PheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeu 104
Db 5204 CTGCGCGCCACCATGTCGATTCGCTGCGCTGCGCTGCATCAGCCGCGCTATGCCC--ACG 5148
QY 105 SerGlnAspLeuAlaIysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
Db 5147 TCGGCCGATTCACCAATCTCCGCCACATCCCGGGACAGATCACCCGGGGATGGTTT 5088
QY 125 AlaAlaAspAlaAlaProPheGlnArgAlaIaIeGluThrIleLeuProAspAspValPro 144
Db 5087 GCCGAGATGCCAGCGCTTTCGCGCGGCTCGCGAGGTTTTCAGCAGCGTACACCG 5028
QY 145 AlaIlePheThrArgGlyGluLeuAlaGlyArg--ArgThrValSerPheAspSerLeu 163
Db 5027 CTGCTCGCGCTCCGCAACTGCGCGGAGATCGTTCCAAATACGTTTCATTTCCAAACCTCG 4968
QY 164 LeuGluGlnProGlyIleGluAlaAspAlaIaPheAlaIaThrIleProAspThr 183
Db 4967 CTTGAAACGAGCCGACCGACGCGGTGATCGTCTTTCATGATGCGGTGCGGCGGATACC 4908
QY 184 IleAlaIysPheLeuPheThrSerGlySerThrIlyLeuProIyAlaIaValProThrThr 203
Db 4907 GTCCGCAATTCCTTTCATGCTACGACGACGATCGTTCGCCAAAGCGTTCATCCAGAG 4848
QY 204 GlnArgMetLeuCysAlaIaasnGlnMetLeuLeuGlnThrPheProValPheGlyGlu 223
Db 4847 CAACGATGTTGTCCTCAATCAGGAAATGATGCGATGTTATGCTTATCTTCCCGAG 4788
QY 224 GluProProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGlySerHisAsn 243
Db 4787 GAGCCGCGCGCTGCTTATATGGCGCCATGAGAACACACGCGCGGGAACAAGCTG 4728
QY 244 IleGlyIleValLeuIyAsnGlyGlyThrTyrTyrLeuAspAspGlyIySProThrAla 263
Db 4727 TTCAACTGCTCTCTATATGCGCGTACTTATATGATGCGCGCAAGCCACACCG 4668
QY 264 GlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrLeuThr 283
Db 4667 GCGCAGATCGCCACACATGCACTGACATCTCCGGACATATCGCCACTGCTATTTCAAC 4608
QY 284 ValProIyGlyTTPGlnIyLeuValIeGlyAlaLeuGluArgAspSerThrLeuArgIu 303
Db 4607 GTTCCGCTGGCCCAAAATGCTGGTCCGACGAGATCGCAAGGACGAGCGCTGTCGCC 4548
QY 304 ArgPhePheAlaArgMetIySLeuPhePhePheAlaAlaAlaGlyLeuSerGlnIyIle 323
Db 4547 AGCTTTTTCGCGACTGAAAGATGCTATGATGCTGCGGTCGCGCATCGCGACACACA 4488
QY 324 TrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGluArgIleArgMetMetAla 343
Db 4487 TGGGATGCGCTGACGAGCTTTCATGACGACAGGTCGCGCATGCGGTCTCATGCGTGC 4428
QY 344 GlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMet 363
Db 4427 GGCCTTCGCTCGACGAGACGCGACCTTCTGCTGCTTTCGACCAACCGACGACGACAAG 4368
QY 364 AlaGlyTyrIleGlyLeuProAlaProGlyCysGluValIyLeuValProValAspGly 383
Db 4367 CCGGCAATATGATGATTCCTCCGCGACGCGCGGACGATGAACCTGCTGCTTCATGGC 4308
QY 384 IySLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTTPArgAlaProGlu 403

gene
CDS
/gene="Atcu1416"
complement(3684..5519)
/gene="Atcu1416"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon start=1
/transl_table=11
/product="long-chain-fatty-acid-CoA-ligase"
/protein_id="AA142422.1"
/db_xref="GI:17739834"
/translacion="MSVKLMSVDPMVEKRPNGEIMWRNDPLGPPQKLANRLHMCCK
LAPRTWMAADGRGRRPMRVSYSABALDRIIRIGQILPHDLSVEEPILVISENSIEMH
LMVLAACHVGTASAITPAYATSAID/TKLADIRGITGCM/PAEDATPFRALGVEVD
DCTPI/VGRNLNEDRSNTPHFETLLETPEVDNDADAVGPDVAKFLFVINGTYY
KAVIOTGRNLSQNMELADCYGFPREPPVVDVAPMHTAAGNFVLVIYNGTYY
IDRGSPFOIQTLDNLNDISPTWFPVPGHEMLVQAMKDEALCSFPRDLKMLM
YAGCAQHTWALTELTSMAITGAHLVAGAGLSTAPSLFCTEPDCKGNGIPI
OGVTKVLDPGRVETLRKGPMTITPGWRNDELTAADDEGFPYRIGTTPVAVADDP
RCGTFEDGMASPNFLQTGTWVAVGPILAQVYMPAGLIRAVITGERRALGALVVP
FLPARELRGQHLSDAEITIRHPSVRQIVAKLSAHQKNSGSHSVRRLVMEADL
REKEBVDKGINORAVLHRLKEVLESLYDGPQVITVGREVA"
5664..6461
/gene="fad"
/note="synonym: Atcu1417"
5664..6461
/gene="fad"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon start=1
/transl_table=11
/product="enoyl-CoA hydratase"
/protein_id="AA142423.1"
/db_xref="GI:17739835"
/translacion="MSEKWTETFFVSYELRCDIALVGLNRPQRNAISDRFEVIAAAAA
AAAEENARVAIRFGHDHFCAGLIDGHIHTPIEGVGSRRMVAFSQIRBHGTIPMW
SLHGAHVCGGIELASTHTRPADEASFPALDEGRGIFGCGGSVRAIRLLGVARS
DMMLGRVASSAEARRNMLVOYAVVRGALAKAQELATTAASNPASLNTAVNALPRIT
QMAEDEDGLVPSFTSSFTPAASPEAEERLRAFLERKRAVRLKAPDTTA"
complement(6511..7431)
/gene="meuf"
/note="synonym: Atcu1418"
complement(6511..7431)
/gene="meuf"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon start=1
/transl_table=11
/product="methylenetetrahydrofolate reductase"
/protein_id="AA142424.1"
/db_xref="GI:17739836"
/translacion="MGSSFWNDNSGADGVVRSNNILAGSIEVMPRTAKIIDPFRNII
LPDGRVALLIDGTPRIADVNSTAKRLHDEGFPVPHLPARGISQKNELEMLRYOOG
EAGVQVALLGGGOKOPAGSPSSIDIMLETGIPDMGPTNHLAIGHREPGANDIDSSG
SRGVQDALLMKDDPASKRIDARLALTQGPAPFAKVIYEAKEIGKMGCTTIPRHVIGAPR
TVLQTLRKALACGVPSLKVQKRPALSRLPVPEPTDFIABLIDHKANPSLIE"
HHLIPGLGIXKPSAMMKKHSATGLANVA"
7739..8473
/gene="Atcu1419"
7739..8473
/gene="Atcu1419"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon start=1
/transl_table=11
/product="transcriptional regulator, GntR family"
/protein_id="AA142425.1"
/db_xref="GI:17739837"
/translacion="MRQVDAATGGRVATELEKILSGELPGKMLFEVSTAEILLDIS
RVALEEDASRLTEBGLNLRDPGGCVVRFGADVDAIEVRGVEGTPAALAESEMIREVVERA
SSLVALEEDATVQOGLDRCGRVDDVDDGVAALNRIITHDOLALCGSEMIREVVERA
SSLPAASBPATLPDPAATGAFRRSLRGAOEOKHVAIYAVREGARAVAREHSRTTA
RNLDEYMRREAPLELAOVPGALISD"
8623..10044
gene

CDs

```

/gene="gcvT"
/note="synonym: Atu1420"
8623. .1004
/gene="gcvT"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="glycine cleavage system T protein,
aminomethyltransferase"
/protein_id="AA42426.1"
/db_xref="GI:17739838"
/translation="MTMAASLTBEVLKQYNPVEMLRNSKIGMYVYPVAGSPNSMRS
EQAMRGAVLFDOSHMDLTYEGPDAEFLLAYHGINSNSNDLNRAKFVPTPPG
HVGADHIFERQDKFLVGRAPTSLSMFCAYGKNVNIKDPSPSEBGRVLR
THRYQYQIGDPDAPIPEKMGNPIDLFKPHVPMINIGSKVVALRGMGAPOLFWK
GPEKDSYILSCILLEAKQAGVNLVRCGSRAYSTNTLESQWIPSLPIGTGGMKD
YRMIDGADSYEATGAIIGDSFVSENIEDYVNPDELGYFIQMKDPIFGKSLAEMK
GKTHRKKTVEENNAEDVYDIAAFRGEEDHYMDIPDPAVYKSTADLILKGGKHVG
MSMFTGVSTVERCLISGLIVNDIQIGDVLTVWGEFDGSAATSTERHKQELRNV
SETPYAAEARNDYADSWRTKAS"

```

Alignment Scores:

Pred. No.:	4,56e-82	Length:	10173
Score:	1294.00	Matches:	266
Percent Similarity:	61.19%	Conservative:	84
Best Local Similarity:	46.50%	Mismatches:	220
Query Match:	41.99%	Indels:	2
DB:	1	Gaps:	2

US-09-750-986D-30 (1-589) x AE009102 (1-10173)

```

QY      5  GlnAlaLeuLeuProPheProGlyArg1LeuGlnArgLeuGlnIstRpAlaLeuThr 24
      :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5444  GATCCGCTTGCGCCCTATCCGCGAGAATGAAATGATGAGCGTCTCTTCACTGTCGTTCG 5385

QY      25  ArgProGlnGlnThrCysValAlaAlaArgAlaAlaAsnGlyLutPArgArgIleSer 44
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5384  GCAACCGAGGCGCATGATGATGGCGGACCGTCAGGGTCGAGCCATGGCCCGGGGTACG 5325

QY      45  TyrAlaGlnUetPheHisAenValArgAlaIleAlaGlnSerLeuLeuProTyArgIleu 64
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5324  TACCCCGAGGCTCTCGCAAGATACCGCGGATCGGCGCAATGCTCTTCTTCGATCATGACCTT 5285

QY      65  SerAlaGlnArgProLeuLeuIleValSerGlyAsnAspLeuGlnIstLeuAla 84
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5264  TCGGTGAGAGGCGCGCTGCGTGGCTTTCGAAATTCATCGAGCATGGGTGATGCGTG 5205

QY      85  PheGlyAlaMetYrAlaGlyIleProTyCysProValSerProAlaTySerLeuLeu 104
      :: :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5204  CTTCGCGCCGACATGTCGGTATCGCTCGCTGCGCATCGCGCGCTTATGCC--ACG 5148

QY      105  SerGlnAspLeuAlaValSerAaGHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5147  TCGGCCGATCTCACCAAACTCGCGGACATCGCGGAGAGATCACCCGGGGAGTGTTTTT 5088

QY      125  AlaAlaAspAlaAlaProPheGlnArgAlaIleGlnThrIleLeuProAspAspValPro 144
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5087  GCCGAGGATCGAGACGCCCTTTCGCGCGGCGCTCGCGCAGAGTTTTCAGACGACGATACCG 5028

QY      145  AlaIlePheThrArgGlyGluLeuAlaGlyArg--ArgThrValSerPheAspSerLeu 163
      :: :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5027  CTGATCGGCTCGGCAACTGCGGAGAGATGTTTCCAAATCGTTTCATTTCCGAACCCCTG 4968

QY      164  LeuGlnGlnProGlyGlyIleGlnAlaAspAsnAlaPheAlaAlaThrArgIleProAspThr 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      4967  CTTGAAACCGAGCCGACCGAAGCGGTTGATCGTCTTTCATGCGGTCGCGGACGGAATACC 4908

QY      184  IleAlaIstPheLeuPheThrSerGlySerThrTyLeuProIstAlaValProThrThr 203
      :: :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      4907  GTGCCCAATTCCTTTTCAAGCTCAGGACACATCGTTCGCGCCAAAGCGGTATTCAGACAG 4848

QY      204  GlnArgMetLeuCysAlaGlnGlnMetLeuLeuGlnIstPheProValPheGlyGlu 223

```

Db 4847 CACGCGATGTTGCTCCATTCAGAAATGATTCGATTTGATGCTACTTTCCGAG 4788
Qy 224 GluProProValLeuValAspTrpLeuProTAsnHisThrPheGlySerHisAsn 243
Db 4787 GACCCCGCGCTCGTGGATTGGCCCATGGAAACAACGCGCGGGAACAGGTG 4728
Qy 244 IleGlyIleValLeuValAsnGlyGlyThrThrThrLeuAspAspGlyLeuProThrAla 263
Db 4727 TTCACCTCGTCTTAATTAAGGGTAACTATTAATCAATCGCGGCAAGCCAGCCG 4668
Qy 264 GlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyThr 283
Db 4667 GCGCAGATGCGCCAGACATCGGACATTCGCGACATTCGCCACCTGGTATTTCAC 4608
Qy 284 ValProIleGlyTrpGluLeuValGlyAlaLeuGluIleArgPheThrLeuArgGlu 303
Db 4607 GTTCCCGCTGGCCAAATGCTGTCAGGCAATCGCAAGGAGGCGCTGCGCGC 4548
Qy 304 ArgPhePheAlaArgMetLeuPhePhePheAlaAlaGlyLeuSerGlnGlyIle 323
Db 4547 AGCTTTTTCGCACTGAAGATGCTGATGATGCCGTGCCGATGGCAGACACA 4488
Qy 324 TrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGluArgIleArgMetMetAla 343
Db 4487 TGGGATGCGCTGACGAGCTTTCATGCGCAACGCTGCCATGCGCTTCATGCTGCC 4428
Qy 344 GlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMet 363
Db 4427 GGCCTCGCTCGACGAGACGCGACCTTCTCGCTGTTGACCGAACCGGACAG 4368
Qy 364 AlaGlyTyThrIleGlyLeuProAlaProGlyCysGluValIleValLeuValProValAspGly 383
Db 4367 CCGGCAATTCGGATCCGCGAGGCGGCTGACGATGAACCTGCGCTTCGATGGC 4308
Qy 384 LeuLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyThrArgAlaProGlu 403
Db 4307 CGGTAGGAATTACGGCTGAAGGCGGACATCAACGCGGTTACTGGCGTAATGGCGAA 4248
Qy 404 GlnAsnAlaGlnAlaPheAspGluGluGlyTyThrCysSerGlyAspAlaIleValLeu 423
Db 4247 TTGACGCGCGCACCTTCGACGAGGAGGTTCTATCGAATCGGCGATACCGTCAATT 4188
Qy 424 AlaAspProAlaAspProGlnIleValLeuMetPheAspGlyArgIleAlaGluAspPhe 443
Db 4187 GCGGTGCGGACATCCGCGCGCTTTTATTTCATGCGCGCATGCGGAAACCTTC 4128
Qy 444 LeuLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
Db 4127 AAGCTGACAGACCGGACCTGGGTGGCGCTTGGCGCGCAATTAGTCAACATG 4068
Qy 464 GlyGlySerTyValLeuAspValValAlaAlaProAspArgGluCysLeuGlyLeu 483
Db 4067 TTCGCGCGGCTGATCCGTATGCCGTATACGAGGCGAAACCGGCGCAATCGGCGCT 4008
Qy 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyValGluAlaSer 503
Db 4007 CTGCTGCTGCATTCATCCGCGCTACGTAGCTGTGTCGCGGTTCGCAACATCTCTCC 3948
Qy 504 AspAlaGlyValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuValArg 523
Db 3947 GATCGGAGATCATCCGCGACCCGAGTGTCAAGGCGCAGATTCGCGGAAGCTTTCGCG 3888
Qy 524 LeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThr 543
Db 3887 CATCAAAACAGCGCGGCTTCGCGCAAGCGCGTCATGCGCATTCGTGATGGAAGAC 3828
Qy 544 ProProSerIleAspIleGlyGluValThrAspIleGlySerIleAsnGlnArgAlaVal 563
Db 3827 GCGCTGCGCTTCGAGAAAGCGAGTTACGACACAGGAGATCAACAGCGCGCGTG 3768
Qy 564 LeuGlnTrpArgSerAlaValAspAlaLeuTy 575

Db 3767 CTTCGATCGAAGAGCTGTAGATCACTTAC 3722

Search completed: February 8, 2005, 02:03:13
Job time : 6717 sec

This Page Blank (uspto)